Technische Universität Braunschweig

Institut Computational Mathematics
AG Partielle Differentialgleichungen

SBP operators for CPR methods

Master's thesis

by

Hendrik Ranocha

Examiner: Prof. Dr. Th. Sonar
Co-examiner: Prof. Dr. H. Löwe

Braunschweig, February 17, 2016

Abstract

Summation-by-parts (SBP) operators have been used in the finite difference framework, providing means to prove conservation and discrete stability by the energy method, predominantly for linear (or linearised) equations. Recently, there have been some approaches to generalise the notion of SBP operators and to apply these ideas to other methods.

The correction procedure via reconstruction (CPR), also known as flux reconstruction (FR) or lifting collocation penalty (LCP), is a unifying framework of high order methods for conservation laws, recovering some discontinuous Galerkin, spectral difference and spectral volume methods.

Using a reformulation of CPR methods relying on SBP operators and simultaneous approximation terms (SATs), conservation and stability are investigated, recovering the linearly stable CPR schemes of Vincent et al. (2011, 2015).

Extensions of SBP methods with diagonal-norm operators to Burgers' equation are possible by a skew-symmetric form and the introduction of additional correction terms.

An analytical setting allowing a generalised notion of SBP methods including modal bases is described and applied to Burgers' equation, resulting in an extension of the previously mentioned skew-symmetric form.

Finally, an extension of the results to multiple space dimensions is presented.

Acknowledgements

First of all, I would like to express my gratitude to Professor Dr. Thomas Sonar, the examiner of this master's thesis. Ever since my first semester, his lectures have been augmenting my interest for mathematics and his advice to intensify my studies by also majoring in mathematics besides physics has encouraged me to take the path I am following now. His everlasting support during my studies includes the possibility to participate in the conference *Recent Developments in the Numerics of Nonlinear Hyperbolic Conservation Laws* at Oberwolfach, where the idea for this master's thesis has developed.

Secondly, I would like to thank Professor Dr. Harald Löwe, the co-examiner of this work. During his lectures, I have awakened to see mathematics as a clear and consequent continuation of physics.

Moreover, I wish to thank Dr. Philipp Öffner, the supervising tutor of this master's thesis, who always had some time for me.

Finally, I thank all the people rendering my life as nice as it is. You know that I am not able to express my feelings that frankly most of the time, but I sincerely hope you feel addressed.

Contents

1 Introduction

Many problems in classical physics can be described by conservation laws, usually formulated as hyperbolic partial differential equations. They appear in fluid mechanics, electrodynamics, space / plasma physics and other areas. Since the problems they describe can be very complicated in nature, obtaining solutions by hand is often impossible. Therefore, the (physical) theory relies on numerical approximations in order to be tested by experiments.

However, hyperbolic conservation laws pose severe problems, even for the mathematical theory of existence and uniqueness of solutions. Up to date, comprehensive results are not available in general. This unsatisfying state of the art is reflected in a lack corresponding results for numerical methods. But without results about convergence, there is no way to know whether differences between numerical solutions and experiments are based on faulty numerical methods or incompleteness of the applied physical theory.

A typical feature of hyperbolic conservation laws is the appearance of shocks, i.e. discontinuities, in finite time. Thus, the classical notion of differentiability is lost and extended notions of the differential equations have to be considered. These possibly lead to different solutions, and new conditions are necessary to single out the physically relevant one – such a solution should exist if the physical theory is adequate to describe the problem at hand. A typical condition derived from physical considerations is an entropy condition. Due to different (sign) conventions, mathematical entropy is convex and should not increase, whereas physical entropy should not decrease.

There are many numerical methods designed for the solution of hyperbolic conservation laws. Besides technical considerations regarding implementation and parallelisation, the important choice of algorithm should be based on desirable properties. For conservation laws, conservativeness and stability of the method are expected. However, as the theory for the continuous problem, discrete stability is hard to achieve in a general way. Therefore, linear stability and nonlinear stability for a specific test problem are considered in this work.

Summation-by-parts (SBP) operators originated in finite difference methods approximately 40 years ago and provide a framework designed for linear (and linearised) problems. However, they have gained a lot of attention in the last years. The *correction procedure via reconstruction* (CPR) is a relatively new framework, unifying several other methods. The aim of this master's thesis is to embed CPR methods in a generalised SBP framework.

Therefore, chapter 2 introduces existing formulations for both SBP operators and CPR methods. Chapter 3 contains the embedding of CPR methods in the framework of (generalised) SBP operators and an investigation of linear stability. Nonlinear stability for Burgers' equation using a skew-symmetric form is investigated in chapter 4. A generalised notion of the methods described hitherto is presented in chapter 5 and extended to multiple dimensions in chapter 6. Finally, the results are summarised in chapter 7 and further directions of research are presented.

2 Existing formulations for SBP operators and CPR methods

In order to fix some notation and introduce the main topics of this work, a brief review of the development of both SBP operators and CPR methods is presented in this chapter. It is not claimed to be a detailed report about the historical evolution of these topics.

This work is concerned with the numerical solution of *partial differential equations* (PDEs), especially hyperbolic conservation laws. Typically, the solution u is sought in some infinite dimensional (real) function space. For numerical approximations, only a finite number of dimensions (often called degrees of freedom) can be considered. Members of these finite dimensional spaces are written underlined, e.g. $\underline{u}$. Linear operators on these finite dimensional spaces are denoted with two underlines, e.g. $\underline{\underline{D}}$.

2.1. Summation-by-parts operators

The development of *summation-by-parts* (SBP) operators has been initiated nearly 40 years ago [Kreiss and Scherer, 1974] in the framework of *finite difference* (FD) methods. The finite dimensional approximation of a function u on an interval $\Omega = [-1, 1] \subset \mathbb{R}$ is given by point values $u(x_i), i \in \{1, \dots, N\}$, where $-1 = x_1 < \cdots < x_N = 1$, i.e. $\underline{u}_i = u(x_i)$. An SBP operator $\underline{\underline{D}}$ for the first derivative fulfils

$$\underline{\underline{M}}\,\underline{\underline{D}} + \underline{\underline{D}}^T\underline{\underline{M}} = \underline{\underline{B}} = \operatorname{diag}(-1, 0, \dots, 0, 1), \tag{2.1}$$

where $\underline{\underline{M}}$ is symmetric positive definite (spd) and approximates the inner product in $L^2(\Omega)$. Therefore, summation-by-parts mimics integration by parts on a discrete level

$$\begin{aligned} \underline{u}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{v} + (\underline{\underline{D}}\,\underline{u})^T\underline{\underline{M}}\,\underline{v} &\approx \int_{-1}^{1} u(x)\,\partial_x v(x)\,\mathrm{d}x + \int_{-1}^{1} \partial_x u(x)\,v(x)\,\mathrm{d}x \\ &= u(x)\,v(x)\Big|_{-1}^{1} = \underline{u}_N\underline{v}_N - \underline{u}_1\underline{v}_1 = \underline{u}^T\underline{\underline{B}}\,\underline{v}. \end{aligned} \tag{2.2}$$

In the following, the dependence on variables as x is not always written explicitly. The measure used in the integrals is an appropriate Lebesgue measure and may be dropped, just as well as the domain of integration.

Traditional FD SBP operators use central differences in the interior of the domain and adapted stencils near the boundaries. As an example, a very simple SBP operator on a uniform grid with mesh spacing Δx is

$$\underline{\underline{D}} = \frac{1}{2\Delta x}\begin{pmatrix} -2 & 2 & 0 & & & \\ -1 & 0 & 1 & 0 & & \\ 0 & -1 & 0 & 1 & 0 & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 0 & 1 \\ & & & 0 & -2 & 2 \end{pmatrix} \tag{2.3}$$

with corresponding diagonal norm matrix

$$\underline{\underline{M}} = \Delta x \ \mathrm{diag}\left(\frac{1}{2}, 1, \ldots, 1, \frac{1}{2}\right). \tag{2.4}$$

Other examples for generalised SBP operators used in this work are given in appendix A.

Just as well-posedness of a PDE depends on boundary conditions, stability of a numerical scheme depends on the way to implement boundary conditions. *Simultaneous approximation terms* (SATs) provide a way to enforce boundary conditions weakly in a stable manner [Carpenter et al., 1994]. The boundary values are treated as unknowns (in contrast to the injection method, where they are inserted at the corresponding nodes) and an SAT is added, driving the solution towards the desired boundary values. As an example, an SAT for the enforcement of the boundary value g_L at the left boundary node x_1 is $\sigma \underline{\underline{M}}^{-1} \underline{e}_1 (\underline{u}_1 - g_L)$, where $\underline{e}_1 = (1, 0, \ldots, 0)^T$ and σ is a real parameter. Therefore, a semidiscretisation of the linear advection equation

$$\partial_t u + \partial_x u = 0 \tag{2.5}$$

in $\Omega = [-1, 1]$ with boundary condition $u(t, -1) = g_L(t)$ and initial condition $u(0, x) = u_0(x)$ can be written

$$\partial_t \underline{u} + \underline{\underline{D}}\, \underline{u} = -\sigma \underline{\underline{M}}^{-1} \underline{e}_1 (\underline{u}_1 - g_L). \tag{2.6}$$

Using the SBP property (2.2) in the energy method, stability of this scheme can be established, see also section 3.3. In this way, SBP operators with SATs are constructed to yield proofs of stability and convergence for linear (or linearised) problems.

In the following, some contributions from several researchers are mentioned that (may) have influenced the author of this master's thesis, before or after developing and writing the main part. Different boundary procedures for SBP operators are compared by Mattsson [2003], resulting in advantages for the SATs. A way to add artificial dissipation using SBP operators is presented by Mattsson et al. [2004]. The advantages of diagonal-norm SBP operators for coordinate transformations are presented by Svärd [2004] and Nordström [2006] for variable coefficients in combination with skew-symmetric formulations. Mattsson and Almquist [2013] used damping at the boundaries as a solution for block-norm SBP operators applied to coordinate transformations / variable coefficients. Connections to quadrature rules have been published by Hicken and Zingg [2013]. Coupling of (nonconforming) block interfaces has been considered by Mattsson and Carpenter [2010], Nissen et al. [2015], Kozdon and Wilcox [2015], Lundquist and Nordström [2015]. Abbas et al. [2009, 2010] and Eriksson et al. [2011] investigated MUSCL schemes of second order in an SBP formulation. SBP operators using extended definitions have been proposed by Mattsson et al. [2014], Fernández et al. [2014a], Fernández and Zingg [2015], Hicken et al. [2015].

Further details and developments can be found in the reviews of Svärd and Nordström [2014], Fernández et al. [2014b], Nordström and Eliasson [2015] and the references cited therein.

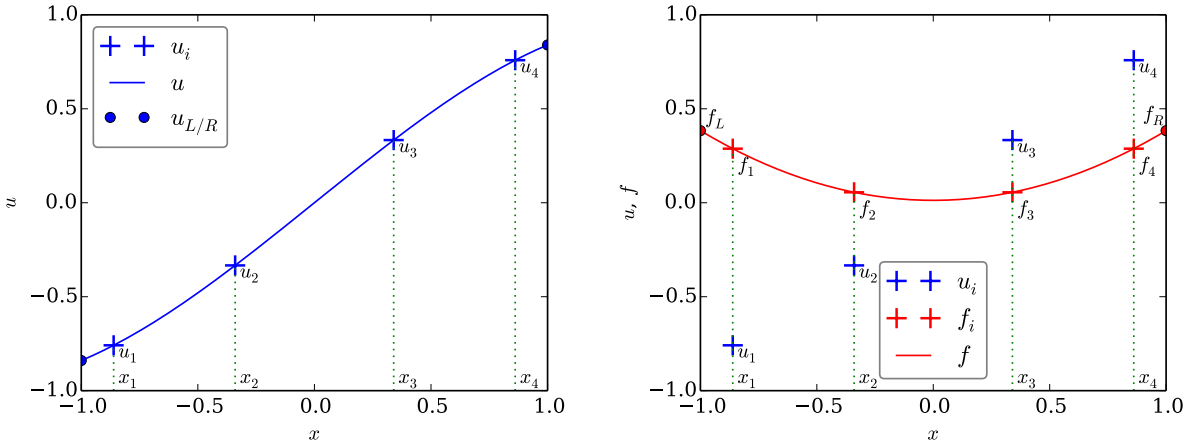
2.2. Correction procedure via reconstruction

The *flux reconstruction* (FR) approach introduced by Huynh [2007] seems to be rather different from finite difference methods. Using special choices of parameters, several other schemes (with according choices of parameters) can be recovered, such as *spectral difference* (SD), *spectral volume* (SV) and *discontinuous Galerkin* (DG). An extension of the flux reconstruction method in one space dimension to triangles has been developed by Wang and Gao [2009] and is known as *lifting collocation penalty* (LCP). Later, Huynh et al. [2014] reviewed these methods and determined the common name *correction procedure via reconstruction* (CPR).

A scalar conservation law

$$\partial_t u + \partial_x f(u) = 0 \tag{2.7}$$

in $\Omega \subset \mathbb{R}$ is approximated numerically by dividing the domain in closed intervals with pairwise disjoint interior and using a nodal polynomial approximation $\underline{u}$ in each cell. Thus, after a mapping to the standard element $[-1,1]$, the coefficients of $\underline{u}$ are the values of u at the nodes $x_0 < \cdots < x_p \in [-1,1]$. The flux $\underline{f}$ is calculated as polynomial interpolation of $f(u)$ at the nodes x_i, i.e. $\underline{f}_i = f(\underline{u}_i) = f(u(x_i))$. By interpolation to the left boundary point -1, the values u_L and f_L are obtained. This procedure is visualised in Figure 2.1 for the flux $f(u) = u^2/2$. The nodes are chosen as Gauß-Legendre nodes in $[-1,1]$ for polynomials of degree $\leq p = 3$.



(a) Function u, $u_i = u(x_i)$.

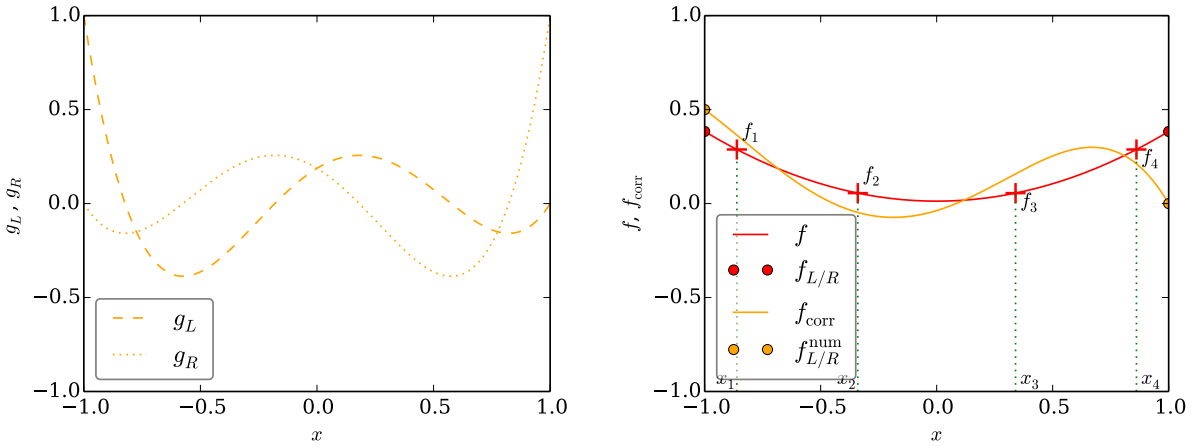
(b) Flux $f(u) = u^2/2$, $f_i = u_i^2/2$.

Figure 2.1.: Visualisation of polynomial collocation used in the *correction procedure via reconstruction*.

In order to get a continuous flux in the whole domain, a common numerical flux f_L^{num} is computed from the boundary values of the neighbouring elements. Using a correction function g_L, fulfilling $g_L(-1) = 1$ and $g_L(1) = 0$ and approximating 0 in $(-1,1)$ in some sense, a correction to the flux given by $\underline{f}$ is $(f_L^{\text{num}} - f_L)g_L$. Using the same procedure at the right boundary 1, the corrected flux is of the form $\underline{f}_{\text{corr}} = \underline{f} + (f_L^{\text{num}} - f_L)\underline{g}_L + (f_R^{\text{num}} - f_R)\underline{g}_R$. Normally, the nodes are chosen symmetric around 0 and $g_L(x) = g_R(-x)$. The correction functions are polynomials of degree $p+1$, i.e. one degree higher than the numerical solution $\underline{u}$. This part of the scheme is visualised in Figure 2.2. The correction function

$$g_L = \frac{(-1)^{p+1}}{2}(\phi_{p+1} - \phi_p) \tag{2.8}$$

is the right Radau polynomial of degree $p+1$ and $g_R(x) = g_L(-x)$. ϕ_i is the Legendre polynomial of degree i, see also appendix A. This choice leads to a *discontinuous Galerkin* (DG) method (with exact mass matrix), as already explained by Huynh [2007]. The numerical fluxes have been chosen arbitrarily as $f_L^{\text{num}}(u_L) = 0.5$, $f_R^{\text{num}}(u_R) = 0$.



(a) Correction functions g_L and g_R.

(b) Corrected / reconstructed flux f_{corr}.

Figure 2.2.: Visualisation of the correction step involved in *flux reconstruction* or *correction procedure via reconstruction*.

Finally, the discrete derivative is evaluated as exact derivative of polynomials, i.e.

$$\partial_t \underline{u} + \underline{\underline{D}}\,\underline{f} + (f_L^{\text{num}} - f_L)\underline{g}'_L + (f_R^{\text{num}} - f_R)\underline{g}'_R = 0. \tag{2.9}$$

The special choice of Lobatto nodes and correction functions named g_2 by Huynh [2007] leads to the *discontinuous Galerkin spectral element method* (DGSEM) with nodal Lobatto-Legendre basis and a lumped mass matrix used by Gassner [2013, 2014], Kopriva and Gassner [2014], Gassner et al. [2016].

In the following, some contributions from several researchers are mentioned that (may) have influenced the author of this master's thesis. Results about linear stability for the advection equation with constant velocity have been obtained by Jameson [2010], Vincent et al. [2011b, 2015]. Vincent et al. [2011a] used von Neumann analysis and extended results about some linearly stable schemes. Connections to other high-order methods are elaborated on by Allaneau and Jameson [2011], Yu and Wang [2013], De Grazia et al. [2014]. First investigations of nonlinear stability have been conducted by Jameson et al. [2012], Witherden and Vincent [2014, 2015]. Some special choices of parameters have been proposed by Asthana and Jameson [2015].

For further details and developments, the review of Huynh et al. [2014] and the references cited therein are recommended.

3 CPR methods using SBP operators

This chapter focuses on a new formulation of CPR methods with special attention paid to SBP operators in one space dimension. Extensions to multiple space dimensions via tensor products are straightforward. Additionally, constant velocity linear advection is used as a test case to investigate linear stability and conservation properties of the schemes.

This chapter has been published by order of Professor Sonar [Ranocha et al., 2015b, 2016].

3.1. The one dimensional setting

After mapping each element to the standard element $[-1,1] \subset \mathbb{R}$, a CPR method can be formulated as

$$\partial_t \underline{u} + \underline{\underline{D}}\, \underline{f} + \underline{\underline{C}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{f}) = 0. \tag{3.1}$$

Here, $\underline{u}, \underline{f}$ are the finite dimensional representation of u, $f(u)$ in the standard element and $\underline{f}^{\text{num}}$ is the representation of the numerical flux on the boundary. The linear operators representing differentiation and restriction (interpolation) to the boundary of the standard element are represented via the matrices $\underline{\underline{D}}$ and $\underline{\underline{R}}$, respectively. Other parameters of the correction operator are encoded in the correction matrix $\underline{\underline{C}}$. Thus, for a given standard element, a CPR method is parametrised by

- A basis $\mathcal{B}$ for the local expansion, determining the derivative and restriction (interpolation) matrices $\underline{\underline{D}}$ and $\underline{\underline{R}}$.
- A correction matrix $\underline{\underline{C}}$, adapted to the chosen basis.

For the representation of an SBP operator, the basis $\mathcal{B}$ has to be associated with a (volume) quadrature rule, given by nodes $z_0, \dots, z_p$ and appropriate positive weights $\omega_0, \dots, \omega_p$. The values of u at the nodes are the coefficients of the local expansion, i.e. $\underline{u} = (u(z_0), \dots, u(z_p))^T$. The quadrature weights determine a positive definite matrix $\underline{\underline{M}} = \operatorname{diag}\left(\omega_0, \dots, \omega_p\right)$ associated with a (discrete) norm $\|u\|_M^2 = \underline{u}^T \underline{\underline{M}}\, \underline{u}$. Besides the volume quadrature rule, there must be a quadrature rule for the boundary, approximating the outward flux through the boundary as in the divergence theorem. In the present one dimensional setting, this quadrature rule is simply given by exact evaluation at the endpoints $\cdot|_{-1}^{1}$. The basis and its associated quadrature rules must satisfy the SBP property

$$\underline{\underline{M}}\, \underline{\underline{D}} + \underline{\underline{D}}^T \underline{\underline{M}} = \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}, \tag{3.2}$$

in order to mimic integration by parts on a discrete level

$$\underline{u}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{v} + (\underline{\underline{D}}\, \underline{u})^T \underline{\underline{M}}\, \underline{v} \approx \int_{-1}^{1} u\, \partial_x v + \int_{-1}^{1} \partial_x u\, v = u\, v \Big|_{-1}^{1} \approx (\underline{\underline{R}}\, \underline{u})^T \underline{\underline{B}}\, (\underline{\underline{R}}\, \underline{u}). \tag{3.3}$$

As an example, consider Gauß-Lobatto-Legendre integration with its associated basis of point values at Lobatto nodes in $[-1,1]$. Then, the restriction and boundary integral matrices reduce to

$$\underline{\underline{R}} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}, \quad \underline{\underline{B}} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.4}$$

Using the special choice $\underline{\underline{C}} = \underline{\underline{M}}^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$ and defining $\underline{\underline{\widetilde{B}}} := \underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}} = \operatorname{diag}(-1, 0, \dots, 0, 1)$, the CPR method of equation (3.1) reduces to

$$\partial_t \underline{u} + \underline{\underline{D}}\,\underline{f} + \underline{\underline{M}}^{-1}\underline{\underline{\widetilde{B}}}\,(\underline{\widetilde{f}}^{\text{num}} - \underline{f}) = 0, \tag{3.5}$$

where $\underline{\widetilde{f}}^{\text{num}} = (f_L^{\text{num}}, 0, \dots, 0, f_R^{\text{num}})$ contains the numerical flux at the left and right boundary and satisfies $\underline{\underline{R}}\,\underline{\widetilde{f}}^{\text{num}} = \underline{f}^{\text{num}}$. Equation (3.5) is the strong form of the DGSEM formulation of Gassner [2013], which he proved to be a diagonal norm SBP operator.

3.2. Conservation

Consider now a CPR method given by a nodal basis of polynomials of degree $\leq p$ and an associated (symmetric) quadrature rule that is exact for polynomials of degree $\leq 2p-1$, for example Gauß-Legendre or Gauß-Lobatto-Legendre quadrature. Then, due to exact integration of polynomials of the form $u\,\partial_x v$, where u, v are polynomials of degree $\leq p$, the SBP property (3.2) automatically holds, see also Kopriva and Gassner [2010]. Let $\underline{1}$ denote the representation of the constant function $x \mapsto 1$ in the chosen basis, i. e. $\underline{1} = (1, \dots, 1)^T$ for a nodal polynomial basis.

In order to investigate conservation properties in the continuous setting, the function u is multiplied with the constant function $x \mapsto 1$ and integrated over the interval (a, b), resulting in

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u = -\int_a^b \partial_x f(u) = -f(u)\Big|_a^b = -\left(f_R - f_L\right). \tag{3.6}$$

Mimicking this derivation in the semidiscrete setting (in the standard element) leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{-1}^{1} u = \frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = -\underline{1}^T\underline{\underline{M}}\left(\underline{\underline{D}}\,\underline{f} + \underline{\underline{C}}\,(\underline{f}^{\text{num}} - \underline{\underline{R}}\,\underline{f})\right). \tag{3.7}$$

Using the SBP property (3.2) results in

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = \underline{1}^T\underline{\underline{D}}^T\underline{\underline{M}}\,\underline{f} - \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{f} - \underline{1}^T\underline{\underline{M}}\,\underline{\underline{C}}\,(\underline{f}^{\text{num}} - \underline{\underline{R}}\,\underline{f}). \tag{3.8}$$

Since discrete differentiation is exact for polynomials of degree $\leq p$ and especially for constant functions, $\underline{\underline{D}}\,\underline{1} = 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = -(\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}} - \underline{1}^T\underline{\underline{M}}\,\underline{\underline{C}})\underline{\underline{R}}\,\underline{f} - \underline{1}^T\underline{\underline{M}}\,\underline{\underline{C}}\,\underline{f}^{\text{num}}. \tag{3.9}$$

Lemma 3.1. *If the assumptions of this subsection are complied with and the correction operator of the CPR method satisfies* $\underline{1}^T\underline{\underline{M}}\,\underline{\underline{C}} = \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}$, *then the scheme is conservative.*

Proof. Inserting the condition into (3.9) gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = -\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\text{num}} = -\left(f_R^{\text{num}} - f_L^{\text{num}}\right), \tag{3.10}$$

due to exact evaluation of the boundary integral for polynomials of degree $\leq p$. Summing up the contributions of all elements and bearing in mind that the numerical flux at the boundary point between two adjacent elements is the same for both, biased only by a factor of -1 for one element but not for the other, results in the global equality

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u = \frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = -\left(f_R - f_L\right) \tag{3.11}$$

also for the numerical scheme. □

Assuming periodic boundary conditions therefore leads to global conservation

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u = \frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = 0. \tag{3.12}$$

Lemma 3.1 proofs conservation across elements. On a sub-element level, conservation for diagonal-norm SBP operators (including boundary nodes) has been proven by Fisher et al. [2013] in the context of the Lax-Wendroff theorem.

3.3. Linear stability

Specializing on a certain type of flux, namely the flux of linear advection with constant velocity 1, the conservation law reduces to

$$\partial_t u + \partial_x u = 0. \tag{3.13}$$

In the continuous setting, proving stability with respect to the L^2 norm is simply an application of integration by parts. Multiplying equation (3.13) by the solution u and integrating over the domain leads to

$$\int u\,\partial_t u = -\int u\,\partial_x u = -\,u^2\Big| + \int \partial_x u\,u. \tag{3.14}$$

Summing up the first and the last equality results in

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_{L^2}^2 = -\,u^2\Big|, \tag{3.15}$$

allowing an estimate of the solution's norm in terms of the initial and boundary conditions, i.e. *well-posedness*. Assuming compact support or periodic boundary conditions simplifies the estimate to $\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_{L^2}^2 = 0$.

Mimicking this manipulations in the discrete setting of an SBP CPR method reads as

$$\int u\,\partial_t u \approx \underline{u}^T\underline{\underline{M}}\,\frac{\mathrm{d}}{\mathrm{d}t}\underline{u} = -\underline{u}^T\underline{\underline{M}}\,\left(\underline{\underline{D}}\,\underline{u} + \underline{\underline{C}}\,(\underline{f}^{\mathrm{num}} - \underline{\underline{R}}\,\underline{u})\right). \tag{3.16}$$

Applying the SBP property (3.2) as in the previous section results in

$$\underline{u}^T\underline{\underline{M}}\,\frac{\mathrm{d}}{\mathrm{d}t}\underline{u} = \underline{u}^T\underline{\underline{D}}^T\underline{\underline{M}}\,\underline{u} - \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u} - \underline{u}^T\underline{\underline{M}}\,\underline{\underline{C}}\,(\underline{f}^{\mathrm{num}} - \underline{\underline{R}}\,\underline{u}). \tag{3.17}$$

Summing up these equations and using the symmetry of the scalar product induced by $\underline{\underline{M}}$ yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = -\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u} - 2\underline{u}^T\underline{\underline{M}}\,\underline{\underline{C}}\,(\underline{f}^{\mathrm{num}} - \underline{\underline{R}}\,\underline{u}). \tag{3.18}$$

Assuming now the special form $\underline{\underline{C}} = \underline{\underline{M}}^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$ simplifies the last equation to

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u} - 2\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\mathrm{num}} = \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u} - 2\underline{f}^{\mathrm{num}}). \tag{3.19}$$

Due to exact evaluation of the boundary terms for $\underline{u}$, representing a polynomial of degree $\leq p$, this can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = u_R(u_R - 2f_R^{\mathrm{num}}) - u_L(u_L - 2f_L^{\mathrm{num}}), \tag{3.20}$$

where the indices R and L indicate values at the right and left boundary, respectively. Therefore, a similar estimate of the norm of the numerical solution in terms of boundary data and the numerical flux is possible. Assuming again periodic boundary conditions or compact support reduces the global rate of change to a sum of local contributions of the form $u_-(u_- - 2f^{\text{num}}) - u_+(u_+ - 2f^{\text{num}})$, where f^{num} is the common numerical flux and u_- is the value u_R on the right boundary of the left element and u_+ is appropriately defined. Using a standard numerical flux of the form

$$f^{\text{num}}(u_-, u_+) = \frac{u_+ + u_-}{2} - \alpha(u_+ - u_-), \tag{3.21}$$

recovering a central scheme for $\alpha = 0$ and a fully upwind scheme for $\alpha = 1$, yields

$$\begin{aligned} &u_-(u_- - 2f^{\text{num}}) - u_+(u_+ - 2f^{\text{num}}) \\ &= u_-^2 - u_+^2 - u_-(u_- + u_+) + 2\alpha u_-(u_+ - u_-) + u_+(u_- + u_+) - 2\alpha u_+(u_+ - u_-) \\ &= -2\alpha(u_+ - u_-)^2. \end{aligned} \tag{3.22}$$

Thus, $\alpha \geq 0$ ensures $\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M \leq 0$ and therefore *stability*, the discrete analogue to well-posedness.

The basic idea of Jameson [2010] to show linear stability is using the equivalence of norms in finite dimensional vector spaces and showing stability not in a regular L^2 norm, but a kind of Sobolev norm involving derivatives, as also explained by Allaneau and Jameson [2011] and used by Vincent et al. [2011b, 2015] to derive linearly stable FR methods. Although the ansatz here is very different, some calculations are similar and in the end, the same schemes will be derived. The difference is, that Vincent et al. used continuous integral norms for their derivations whereas this setting uses fully discrete norms adapted to the solution point coordinates. Therefore, they could not recognize any influence of the solution points on the stability properties in the linear case. For the nonlinear case, the influence of these nodes was stressed by Jameson et al. [2012].

Following these ideas, stability is investigated for a discrete norm given by $\underline{\underline{M}} + \underline{\underline{K}}$, where $\underline{\underline{M}}$ is the matrix associated as usual with the quadrature rule given by the polynomial basis and $\underline{\underline{K}}$ is a symmetric matrix satisfying $\underline{\underline{M}} + \underline{\underline{K}} > 0$, i.e. positive definite. Then, the rate of change of the discrete norm $\|u\|_{M+K}^2 = \underline{u}^T(\underline{\underline{M}} + \underline{\underline{K}})\underline{u}$ can be computed via

$$\underline{u}^T(\underline{\underline{M}} + \underline{\underline{K}})\frac{\mathrm{d}}{\mathrm{d}t}\underline{u} = -\underline{u}^T(\underline{\underline{M}} + \underline{\underline{K}})\left(\underline{\underline{D}}\,\underline{u} + \underline{\underline{C}}\,(\underline{f}^{\text{num}} - \underline{\underline{R}}\,\underline{u})\right), \tag{3.23}$$

which can also be written as

$$\begin{aligned} \underline{u}^T(\underline{\underline{M}} + \underline{\underline{K}})\tfrac{\mathrm{d}}{\mathrm{d}t}\underline{u} = &-\underline{u}^T\underline{\underline{K}}\,\underline{\underline{D}}\,\underline{u} - \underline{u}^T(\underline{\underline{M}} + \underline{\underline{K}})\underline{\underline{C}}\,(\underline{f}^{\text{num}} - \underline{\underline{R}}\,\underline{u}) \\ &+ \underline{u}^T\underline{\underline{D}}^T\underline{\underline{M}}\,\underline{u} - \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u}, \end{aligned} \tag{3.24}$$

due to the SBP property (3.2). Again, adding the last two equations yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_{M+K}^2 = -2\underline{u}^T\underline{\underline{K}}\,\underline{\underline{D}}\,\underline{u} - 2\underline{u}^T(\underline{\underline{M}} + \underline{\underline{K}})\underline{\underline{C}}\,(\underline{f}^{\text{num}} - \underline{\underline{R}}\,\underline{u}) - \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u}. \tag{3.25}$$

The last term contains only boundary values and is thus unproblematic. The second term can be rendered as a boundary term by enforcing the correction matrix to be $\underline{\underline{C}} = (\underline{\underline{M}} + \underline{\underline{K}})^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$, analogously to the previous procedure. Then, the only term remaining to be estimated is the first one.

In the following sections, the multiple parameter family of FR methods of Vincent et al. [2015] will be reconsidered using the view of SBP operators. These parameters force the first term to vanish, because $\underline{\underline{K}}\,\underline{\underline{D}}$ is chosen to be antisymmetric. Then, using $\underline{\underline{C}} = (\underline{\underline{M}} + \underline{\underline{K}})^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$, the last equation can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u\|^2_{M+K} = -2\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{f}^{\text{num}} - \underline{\underline{R}}\,\underline{u}) - \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u} = \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u} - 2\underline{f}^{\text{num}}), \tag{3.26}$$

allowing the same estimates as before, leading to linear stability. This proves the following

Lemma 3.2 (see also Vincent et al. [2015, Thm.1]). *If the SBP CPR method is given by* $\underline{\underline{C}} = (\underline{\underline{M}} + \underline{\underline{K}})^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$, *where* $\underline{\underline{M}} + \underline{\underline{K}}$ *is positive definite and* $\underline{\underline{K}}\,\underline{\underline{D}}$ *is antisymmetric, then the method is linearly stable in the discrete norm* $\|\cdot\|_{M+K}$ *induced by* $\underline{\underline{M}} + \underline{\underline{K}}$.

3.4. Symmetry

In order to recognize the FR method associated with an SBP CPR method, it suffices to identify the correction matrix $\underline{\underline{C}}$ with the derivatives of the left and right correction function g_L, g_R. Using again a nodal polynomial basis with symmetric nodes $\xi_0, \dots, \xi_p$ in the standard element, writing

$$\underline{\underline{C}} = \begin{pmatrix} g_L'(\xi_0) & g_R'(\xi_0) \\ \vdots & \vdots \\ g_L'(\xi_p) & g_R'(\xi_p) \end{pmatrix} \tag{3.27}$$

provides the required identification of SBP CPR parameters and FR correction functions. Note that $g_L(-1) = 1 = g_R(1)$ is required, so that the integration constant is fixed. The symmetry property $g_R(\xi) = g_L(-\xi)$ (and therefore also $g_R'(\xi) = -g_L'(-\xi)$) should be satisfied for the correction procedure in order not to get any bias to one direction. Translated to the CPR method, this requires

$$\underline{\underline{C}} = \begin{pmatrix} g'(\xi_0) & -g'(\xi_p) \\ \vdots & \vdots \\ g'(\xi_p) & -g'(\xi_0) \end{pmatrix}, \tag{3.28}$$

dropping the index for g_L and using the symmetry of g_L, g_R, and the nodes $\xi_0, \dots, \xi_p$.

Assume that the nodal basis is associated with a symmetric quadrature that is exact for polynomials of degree $\leq p$. Then, a coordinate transformation to Legendre polynomials, i.e. from a nodal basis to a modal basis, is given by the Vandermonde matrix $\underline{\underline{V}}$ with $V_{i,j} = \phi_j(\xi_i)$, where $\phi_j, j = 0, \dots, p$ are the Legendre polynomials, see also appendix A. Writing matrices and vectors with respect to the Legendre basis using $\hat{\cdot}$, the transformation is $\underline{\underline{V}}\,\underline{\hat{u}} = \underline{u}$. Therefore, the operator matrices like the derivative matrix transform according to $\underline{\underline{\hat{D}}} = \underline{\underline{V}}^{-1}\underline{\underline{D}}\,\underline{\underline{V}}$ and the matrices associated with bilinear forms like $\underline{\underline{M}}$ and $\underline{\underline{K}}$ can be computed as $\underline{\underline{\hat{M}}} = \underline{\underline{V}}^T\underline{\underline{M}}\,\underline{\underline{V}}$.

Because the transformation from Lagrange to Legendre polynomials does not change the basis of the boundary, which is still a nodal basis for a quadrature (indeed, in this one dimensional setting, it is an exact evaluation), the modal correction matrix is $\underline{\underline{\hat{C}}} = \underline{\underline{V}}^{-1}\underline{\underline{C}}$, i.e.

$$\underline{\underline{\hat{C}}} = \underline{\underline{V}}^{-1}\underline{\underline{C}} = \underline{\underline{V}}^{-1}(\underline{g_L'}, \underline{g_R'}) = (\underline{\hat{g_L'}}, \underline{\hat{g_R'}}). \tag{3.29}$$

Because of the alternating symmetry and antisymmetry of the Legendre polynomials, the symmetry condition (3.28) is translated to

$$\underline{\underline{\hat{C}}} = (\hat{\underline{g_L}}, \hat{\underline{g_R}}) = \begin{pmatrix} -c_o & c_0 \\ c_1 & c_1 \\ \vdots & \vdots \\ (-1)^{p+1}c_p & c_p \end{pmatrix}, \tag{3.30}$$

for some coefficients $c_0, \dots, c_p$. Using $\underline{\underline{\hat{C}}} = \underline{\underline{V}}^{-1}\underline{\underline{C}}$, this becomes

$$\begin{aligned} \begin{pmatrix} -c_0 & c_0 \\ c_1 & c_1 \\ \vdots & \vdots \\ (-1)^{p+1}c_p & c_p \end{pmatrix} &= \underline{\underline{\hat{C}}} = \underline{\underline{V}}^{-1}\underline{\underline{C}} \\ &= \underline{\underline{V}}^{-1}(\underline{\underline{M}} + \underline{\underline{K}})^{-1}\underline{\underline{R}}^T\underline{\underline{B}} = \underline{\underline{V}}^{-1}(\underline{\underline{M}} + \underline{\underline{K}})^{-1}\underline{\underline{V}}^{-T}\underline{\underline{V}}^T\underline{\underline{R}}^T\underline{\underline{B}} \\ &= (\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}})^{-1}\underline{\underline{\hat{R}}}^T\underline{\underline{B}}. \end{aligned} \tag{3.31}$$

The modal restriction matrix is given by the values of the Legendre polynomials ϕ_i at -1 and 1, i.e. $\phi_i(-1) = (-1)^i$ and $\phi_i(1) = 1$. Therefore, the symmetry condition reduces to

$$\begin{pmatrix} -c_0 & c_0 \\ c_1 & c_1 \\ \vdots & \vdots \\ (-1)^{p+1}c_p & c_p \end{pmatrix} = (\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}})^{-1} \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ \vdots & \vdots \\ (-1)^p & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = (\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}})^{-1} \begin{pmatrix} -1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ (-1)^{p+1} & 1 \end{pmatrix} \tag{3.32}$$

A sufficient condition for this equality in analogy to Vincent et al. [2015, Thm. 2] is given by

Lemma 3.3. *If for* $\underline{\underline{\hat{J}}} = \operatorname{diag}\left(-1, 1, \dots, (-1)^{p+1}\right)$ *the condition*

$$\underline{\underline{\hat{J}}}\,(\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}}) = (\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}})\underline{\underline{\hat{J}}} \tag{3.33}$$

is satisfied, then the SBP CPR method is symmetric in the sense of equation (3.28).

Proof. Comparing the rows of equation (3.32) leads to the conditions

$$\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_p \end{pmatrix} = (\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}})^{-1} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{3.34}$$

$$\underline{\underline{\hat{J}}} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_p \end{pmatrix} = \begin{pmatrix} -c_0 \\ c_1 \\ \vdots \\ (-1)^{p+1}c_p \end{pmatrix} = (\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}})^{-1} \begin{pmatrix} -1 \\ 1 \\ \vdots \\ (-1)^{p+1} \end{pmatrix} = (\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}})^{-1}\underline{\underline{\hat{J}}} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{3.35}$$

The first condition determines the coefficients $c_0, \dots, c_p$ and the second one is automatically satisfied if $(\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}})$ and $\underline{\underline{\hat{J}}}$ commute. □

If the quadrature is exact of order $2p-1$, $\underline{\underline{\hat{M}}}$ is still a diagonal matrix, because the Legendre polynomials are orthogonal. The entries with index 0 to $p-1$ are the correct norms of the corresponding Legendre polynomials and the last entry may be changed. For Gauß-Lobatto-Legendre quadrature, the last entry is $\underline{\underline{\hat{M}}}_{p,p} = \frac{2}{p}$, as used by Gassner and Kopriva [2011]. For Gauß-Legendre quadrature, the last entry is the correct value $\frac{2}{2p-1}$, because the quadrature is exact for polynomials of degree $\leq 2p+1$, see also equation (3.47) in section 3.6. In this case a new result shows that the condition of Lemma 3.3 is automatically satisfied:

Lemma 3.4. *If the SBP CPR method is associated with a quadrature of order $2p-1$, $\underline{\underline{K}}\,\underline{\underline{D}} + \underline{\underline{D}}^T\underline{\underline{K}}^T = 0$ and $\underline{\underline{M}} + \underline{\underline{K}}$ is positive definite ($\underline{\underline{M}}$ is positive definite by definition), then the symmetry condition $\underline{\hat{J}}\,(\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}}) = (\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}})\underline{\hat{J}}$ of Lemma 3.3 is satisfied.*

Proof. Because the quadrature is exact for polynomials of order $\leq 2p-1$, the modal mass matrix $\underline{\underline{\hat{M}}}$ is diagonal and commutes with $\underline{\hat{J}}$. Therefore, it suffices to prove the commutativity with $\underline{\underline{\hat{K}}}$.

In the following part of the proof, the notation using $\hat{\cdot}$ and $\underline{\cdot}$ is dropped due to simplicity.

Proof for $JK = KJ$ by induction on p: For $p = 1$, the relevant matrices are

$$J = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad K = \begin{pmatrix} k_{00} & k_{01} \\ k_{01} & k_{11} \end{pmatrix}. \tag{3.36}$$

Therefore, $KD + D^TK = 0$ implies

$$\begin{pmatrix} 0 & k_{00} \\ 0 & k_{01} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ k_{00} & k_{01} \end{pmatrix} = 0, \tag{3.37}$$

i.e. $k_{00} = k_{01} = 0$. Using this results in $JKJ = K$.

$p \to p+1$: $J_+K_+J_+ = K_+$ has to be proven using the result for K, J. The matrices are given by

$$J_+ = \begin{pmatrix} J & 0 \\ 0 & (-1)^p \end{pmatrix}, \quad D_+ = \begin{pmatrix} D & d \\ 0 & 0 \end{pmatrix}, \quad K_+ = \begin{pmatrix} K & k \\ k^T & \kappa \end{pmatrix}. \tag{3.38}$$

Using induction, the equation is

$$J_+K_+J_+ = J_+\begin{pmatrix} KJ & (-1)^pk \\ k^TJ & (-1)^p\kappa \end{pmatrix} = \begin{pmatrix} JKJ & (-1)^pJk \\ (-1)^pk^TJ & (-1)^{2p}\kappa \end{pmatrix} = \begin{pmatrix} K & (-1)^pJk \\ (-1)^pk^TJ & \kappa \end{pmatrix}. \tag{3.39}$$

Thus, it remains to show $k = (-1)^pJk$ using $D_+^TK_+ + K_+D_+ = 0$, i.e.

$$\begin{pmatrix} D^TK & D^Tk \\ d^TK & d^Tk \end{pmatrix} + \begin{pmatrix} KD & Kd \\ k^TD & k^Td \end{pmatrix} = \begin{pmatrix} D^TK + KD & D^Tk + Kd \\ d^TK + k^TD & 2d^Tk \end{pmatrix} = 0. \tag{3.40}$$

$k = (-1)^pJk$ can be written as $\operatorname{diag}\big(1 + (-1)^p, \ldots, 2, 0, 2\big)\, k = 0$. That is, $k_p = k_{p-2} = \cdots = 0$ is to be proven.

Calculating the product of J and D results in

$$\begin{aligned} JDJ &= \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & -1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & \dots \\ 0 & 0 & 3 & 0 & 3 & \dots \\ 0 & 0 & 0 & 5 & 0 & \dots \\ 0 & 0 & 0 & 0 & 7 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & -1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & \dots \\ 0 & 0 & -3 & 0 & -3 & \dots \\ 0 & 0 & 0 & 5 & 0 & \dots \\ 0 & 0 & 0 & 0 & -7 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & -1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = -D. \end{aligned} \tag{3.41}$$

Using $J^2 = I$, $J^T = J$, and $D^T J = -JD^T$, yields

$$D^T(I - (-1)^p J)k = D^T k - (-1)^p D^T Jk = D^T k + (-1)^p JD^T k. \tag{3.42}$$

Using equation (3.40) together with $JK = KJ$ results in

$$D^T(I - (-1)^p J)k = (I + (-1)^p J)D^T k = -(I + (-1)^p J)Kd = -K(I + (-1)^p J)d. \tag{3.43}$$

Since $(I + (-1)^p J) = \operatorname{diag}(\dots, 0, 2, 0)$ and $d = (\dots, *, 0, *)^T$, $(I + (-1)^p J)d = 0$ and therefore also $D^T(I - (-1)^p J)k = 0$. Here and in the following, $*$ is a placeholder for an arbitrary real number. Because of the implication

$$0 = D^T x = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & \dots \\ 0 & 3 & 0 & 0 & \dots \\ 1 & 0 & 5 & 0 & \dots \\ 0 & 3 & 0 & 7 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} x_0 \\ \vdots \\ x_p \end{pmatrix} \implies x = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x_p \end{pmatrix} \tag{3.44}$$

and $I - (-1)^p J = \operatorname{diag}(\dots, 2, 0, 2, 0, 2)$, one can deduce that $k = (\dots, 0, *, 0, *, *)^T$. Finally, using $d^T k = 0$ from equation (3.40) yields $k_p = k_{p-2} = \dots = 0$ and finishes the proof. □

3.5. Summary

The results of the previous sections are summed up in the following

Theorem 3.5. *Let a one dimensional CPR method be given by a nodal basis $\mathcal{B}$ of polynomials of degree $\leq p$, associated with a quadrature, given by symmetric nodes $z_0, \dots, z_p \in [-1, 1]$ and positive weights $\omega_0, \dots, \omega_p > 0$, that is exact for polynomials of degree $\leq 2p - 1$. Let*

- $\underline{\underline{M}} = \operatorname{diag}\left(\omega_0, \dots, \omega_p\right) > 0$ *be the (positive definite and diagonal) mass matrix associated with a bilinear volume quadrature,*

- $\underline{\underline{R}}$ *be the restriction operator, performing an interpolation to the boundary,*
- $\underline{\underline{B}} = \operatorname{diag}(-1, 1)$ *be the boundary matrix, associated with an integral along the outer normal of the boundary, and*
- $\underline{\underline{D}}$ *be the discrete derivative matrix, associated with the divergence operator,*

satisfying the SBP property $\underline{\underline{M}}\,\underline{\underline{D}} + \underline{\underline{D}}^T\underline{\underline{M}} = \underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}$. *Then the following results are valid:*

1. *If* $\underline{1}^T\underline{\underline{M}}\,\underline{\underline{C}} = \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}$, *where* $\underline{1}$ *is the representation of the constant function* $\xi \mapsto 1$, *then the SBP CPR method with correction parameters* $\underline{\underline{C}}$ *is conservative (see Lemma 3.1).*
2. *If* $\underline{\underline{C}} = (\underline{\underline{M}} + \underline{\underline{K}})^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$, *where* $\underline{\underline{M}} + \underline{\underline{K}}$ *is positive definite and* $\underline{\underline{K}}\,\underline{\underline{D}}$ *is antisymmetric, then the SBP CPR method given by* $\underline{\underline{C}}$ *is linearly stable in the discrete norm* $\|\cdot\|_{M+K}$ *induced by* $\underline{\underline{M}} + \underline{\underline{K}}$ *(see Lemma 3.2).*
3. *If again* $\underline{\underline{C}} = (\underline{\underline{M}} + \underline{\underline{K}})^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$, *where* $\underline{\underline{M}} + \underline{\underline{K}}$ *is positive definite and* $\underline{\underline{K}}\,\underline{\underline{D}}$ *is antisymmetric, then the SBP CPR method given by* $\underline{\underline{C}}$ *is associated with a FR scheme using symmetric correction functions* $g_L(\xi) = g_R(-\xi)$ *(see Lemmata 3.3 and 3.4).*

3.6. The one parameter family of Vincent et al. (2011)

The approach of Vincent et al. [2011b] can be formulated as enforcing $\underline{\underline{K}}\,\underline{\underline{D}} = 0$ by setting $\underline{\underline{K}} = c(\underline{\underline{D}}^p)^T\underline{\underline{D}}^p$, because $\underline{\underline{D}}^{p+1} = 0$ (polynomials of degree $\leq p$). However, in this work the ansatz $\underline{\underline{K}} = \kappa(\underline{\underline{D}}^p)^T\underline{\underline{M}}\,\underline{\underline{D}}^p$ is chosen to allow an interpretation in terms of discrete norms. Additionally, the transformation of the matrices during a change of the basis is only handled consistently in this way. In the following section, Gauß- and Lobatto-Legendre quadrature rules accompanied by the associated nodal polynomial basis of degree $\leq p$ are considered. Therefore, the leading assumptions of Theorem 3.5 are satisfied.

For concrete computations, again a change to the Legendre basis is advantageous. In these coordinates, the derivative matrix to the power of p is simply (the p-th derivative of a polynomial of degree $\leq p-1$ is identically zero and $p!$ times its leading coefficient for a polynomial of degree p)

$$\hat{\underline{\underline{D}}}^p = \begin{pmatrix} 0 & \dots & 0 & p!\,a_p \\ 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad a_p = \frac{(2p)!}{2^p(p!)^2}, \tag{3.45}$$

referring to the leading coefficient of the Legendre polynomial of degree p in the same way as Vincent et al. [2011b] as a_p. Therefore, using $\hat{\underline{\underline{M}}} = \operatorname{diag}(2, *, \dots, *)$, the ansatz for $\hat{\underline{\underline{K}}}$ becomes

$$\hat{\underline{\underline{K}}} = \kappa \begin{pmatrix} 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 2a_p^2(p!)^2 \end{pmatrix}. \tag{3.46}$$

The choice of the basis influences further computations through the mass matrix $\hat{\underline{\underline{M}}}$. In the following, variables associated with the Gauß-Legendre and Lobatto-Legendre basis are denoted using

a superscript G and L, respectively. Gauß-Legendre quadrature is exact for polynomials of degree $\leq 2p+1$ and Lobatto-Legendre quadrature is exact for polynomials of degree $\leq 2p-1$. Therefore, $\underline{\underline{\hat{M}}}^G$ and $\underline{\underline{\hat{M}}}^L$ are both diagonal and the last entry of $\underline{\underline{\hat{M}}}^L$ is $\frac{2}{p}$ in accordance with Gassner and Kopriva [2011]:

$$\underline{\underline{\hat{M}}}^G = \operatorname{diag}\left(2, \frac{2}{3}, \dots, \frac{2}{2p-1}, \frac{2}{2p+1}\right), \quad \underline{\underline{\hat{M}}}^L = \operatorname{diag}\left(2, \frac{2}{3}, \dots, \frac{2}{2p-1}, \frac{2}{p}\right). \tag{3.47}$$

Therefore, $\underline{\underline{M}}^G + \underline{\underline{K}}$ and $\underline{\underline{M}}^L + \underline{\underline{K}}$ are positive definite if and only if

$$\kappa > \kappa_-^G := -\frac{1}{(2p+1)a_p^2(p!)^2}, \quad \kappa > \kappa_-^L := -\frac{1}{pa_p^2(p!)^2}, \tag{3.48}$$

respectively. Therefore, the associated SBP CPR methods given by $\underline{\underline{C}} = (\underline{\underline{M}} + \underline{\underline{K}})^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$ are linearly stable and conservative by Theorem 3.5 if κ is chosen accordingly to (3.48). In addition, they are conservative, since

$$\begin{aligned}\underline{\hat{1}}^T\underline{\underline{\hat{M}}}\,\underline{\underline{\hat{C}}} &= \underline{\hat{1}}^T\underline{\underline{\hat{M}}}\,(\underline{\underline{\hat{M}}} + \underline{\underline{\hat{K}}})^{-1}\underline{\underline{\hat{R}}}^T\underline{\underline{\hat{B}}} \\ &= (1,0,\dots,0)\operatorname{diag}(2,*,\dots,*)\operatorname{diag}\left(\frac{1}{2},*,\dots,*\right)\underline{\underline{\hat{R}}}^T\underline{\underline{\hat{B}}} = \underline{\hat{1}}^T\underline{\underline{\hat{R}}}^T\underline{\underline{\hat{B}}}.\end{aligned} \tag{3.49}$$

To compare the resulting methods with the ones obtained by Vincent et al. [2011b], equation (3.29) can be used. To compute $\underline{\underline{C}}$ explicitly, the restriction matrix $\underline{\underline{R}}$ has to be computed in the Legendre basis. Describing interpolation to the boundary, using $\phi_i(1) = 1$ and $\phi_i(-1) = (-1)^i$ it can be written as

$$\underline{\underline{\hat{R}}} = \begin{pmatrix} 1 & -1 & 1 & \dots & (-1)^p \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}. \tag{3.50}$$

Therefore, computing $\underline{\underline{C}} = (\underline{\underline{M}} + \underline{\underline{K}})^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$ explicitly results in

$$\begin{aligned}\underline{\underline{\hat{C}}}^{G/L} &= \operatorname{diag}\left(\frac{1}{2}, \dots, \frac{2}{2p-1}, *^{G/L}\right)\begin{pmatrix} -1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ (-1)^{p+1} & 1 \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{3}{2} & \frac{3}{2} \\ \vdots & \vdots \\ (-1)^p\frac{2p-1}{2} & \frac{2p-1}{2} \\ (-1)^{p+1}*^{G/L} & *^{G/L} \end{pmatrix} = (\underline{\hat{g}_L}^{G/L}, \underline{\hat{g}_R}^{G/L}),\end{aligned} \tag{3.51}$$

where $*^G = \left(\frac{2}{2p+1} + 2\kappa a_p^2(p!)^2\right)^{-1}$ and $*^L = \left(\frac{2}{p} + 2\kappa a_p^2(p!)^2\right)^{-1}$. The (symmetric) correction functions of Vincent et al. [2011b] are given by

$$g_L = \frac{(-1)^p}{2}\left(\phi_p - \frac{\eta_p\phi_{p-1} + \phi_{p+1}}{1+\eta_p}\right), \quad g_R = \frac{1}{2}\left(\phi_p + \frac{\eta_p\phi_{p-1} + \phi_{p+1}}{1+\eta_p}\right), \tag{3.52}$$

where

$$\eta_p = c\frac{2p+1}{2}a_p^2(p!)^2. \tag{3.53}$$

Therefore, in order to compare the results, it remains to compute the derivatives of (3.52). The derivative matrix of size $(p+2)\times(p+2)$ for even p is given as

$$\underline{\underline{\hat{D}}}_+ = \begin{pmatrix} 0 & 1 & 0 & 1 & \dots & 1 & 0 & 1 \\ & 0 & 3 & 0 & \dots & 0 & 3 & 0 \\ & & 0 & 5 & \dots & 5 & 0 & 5 \\ & & & \ddots & \dots & \vdots & \vdots & \vdots \\ & & & & & & 0 & 2p+1 \\ & & & & & & & 0 \end{pmatrix} \tag{3.54}$$

and as

$$\underline{\underline{\hat{D}}}_+ = \begin{pmatrix} 0 & 1 & 0 & 1 & \dots & 1 & 0 \\ & 0 & 3 & 0 & \dots & 0 & 3 \\ & & 0 & 5 & \dots & 5 & 0 \\ & & & \ddots & \dots & \vdots & \vdots \\ & & & & & 0 & 2p+1 \\ & & & & & & 0 \end{pmatrix} \tag{3.55}$$

for odd p. Multiplication with

$$\underline{\hat{g_L}} = \frac{(-1)^p}{2}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\frac{\eta_p}{1+\eta_p} \\ 1 \\ -\frac{1}{1+\eta_p} \end{pmatrix} \tag{3.56}$$

results for both even and odd p in the same coefficients with indices 0 to $p-1$ as in (3.51) and thus, in order to get the same methods, the last coefficient has to be the same, resulting in the equation

$$*^{G/L} = \frac{2p+1}{2}\frac{1}{1+\eta_p} = \frac{2p+1}{2}\frac{1}{1+c\frac{2p+1}{2}a_p^2(p!)^2}. \tag{3.57}$$

Inserting $*^{G/L}$ from above, this results in

$$(*^G)^{-1} = \frac{2}{2p+1} + 2\kappa^G a_p^2(p!)^2 = \frac{2}{2p+1} + ca_p^2(p!)^2 \tag{3.58}$$

and

$$(*^L)^{-1} = \frac{2}{p} + 2\kappa^L a_p^2(p!)^2 = \frac{2}{2p+1} + ca_p^2(p!)^2, \tag{3.59}$$

respectively, and therefore the parameter c of Vincent et al. [2011b] can be expressed as

$$c = 2\kappa^G = 2\kappa^L + c_{Hu}, \quad c_{Hu} = 2\frac{p+1}{p(2p+1)a_p^2(p!)^2}, \tag{3.60}$$

where c_{Hu} is the coefficient recovering the scheme named g_2 by Huynh [2007]. This scheme is exactly the same as the DGSEM scheme with a nodal basis at Lobatto-Legendre nodes and a lumped mass matrix used by Gassner [2013, 2014], Kopriva and Gassner [2014], Gassner et al. [2016] and proven to be an SBP scheme.

3.7. The multi parameter family of Vincent et al. (2015)

The results for the multi parameter family of linearly stable and conservative schemes of Vincent et al. [2015] are similar to those obtained in the previous section about the one parameter family – as expected, since the one parameter family is contained in the extended range of schemes.

The calculations of Vincent et al. [2015] used an exact mass matrix (in the Legendre basis) and are thus valid for Gauß-Legendre points. Using Lobatto-Legendre quadrature will result in a transformed parameter space, recovering the same schemes as before, similar to the previous section. In contrast to their results, the solution point coordinates are considered to be an important parameter of an SBP CPR method and thus included in the analysis. Therefore, discrete norms are investigated and stability results are stated in these discrete norms.

3.8. Numerical examples

In order validate the implementation, the numerical experiments presented by Vincent et al. [2011b] are repeated. The conservation law solved is the linear advection equation (3.13) with constant velocity 1 in one space dimension in the interval $[-1, 1]$ with periodic boundary conditions. The initial condition is

$$u_0(x) = e^{-20x^2}. \tag{3.61}$$

Several SBP CPR methods with $N = 10$ equally spaced elements of order $p = 3$ are utilised as semidiscretisations and the classical fourth order Runge-Kutta method with $50,000$ steps is used to obtain the solution in the time interval $[0, 20]$, i.e. ten traversals of the initial data are regarded.

Results for a Lobatto-Legendre basis and the central numerical flux are shown in Figures 3.1, 3.2 and 3.3. Four different values of the parameter c for the correction matrix $\underline{\underline{C}}$ are used, the same as presented by Vincent et al. [2011b]: $c = c_-/2 = \frac{-1}{(2p+1)(a_p\, p!)^2} < 0$ (a negative parameter near the boundary value c_- for stable schemes), $c = c_0 = 0$ (no additional matrix $\underline{\underline{K}}$ for exact integration, i.e. in the framework of Vincent et al. [2011b], corresponding to Gauß-Legendre points in the framework presented here), $c = c_{SD} = \frac{2p}{(2p+1)(p+1)(a_p\, p!)^2}$ (recovering a spectral difference method), and $c = c_{Hu} = \frac{2(p+1)}{(2p+1)p(a_p\, p!)^2}$ (using the correction functions named g_2 by Huynh [2007], corresponding to the DGSEM of Gassner [2013]). Figure 3.1 consists of plots of the solution at $t = 20$ (in blue) and the initial profile at $t = 0$ (in green). In Figure 3.2, the squared L_2 norms computed via Gauß (blue) and Lobatto (green) quadrature in the time interval $[0, 20]$ are plotted. Finally, Figure 3.3 provides a zoomed in view of the time interval $[0, 0.8]$. The solutions are visually the same as those obtained by Vincent et al. [2011b]. Since c_{Hu} corresponds to the correction function g_2 of Huynh [2007] and the corresponding SBP CPR method is the same as the DGSEM of Gassner [2013], the energy computed via Lobatto quadrature remains constant for this choice of c. The results obtained by using a Gauß-Legendre basis look very much the same at this resolution and are consequently not printed.

In the CPR framework, the solution is approximated as a piecewise polynomial function. Thus,

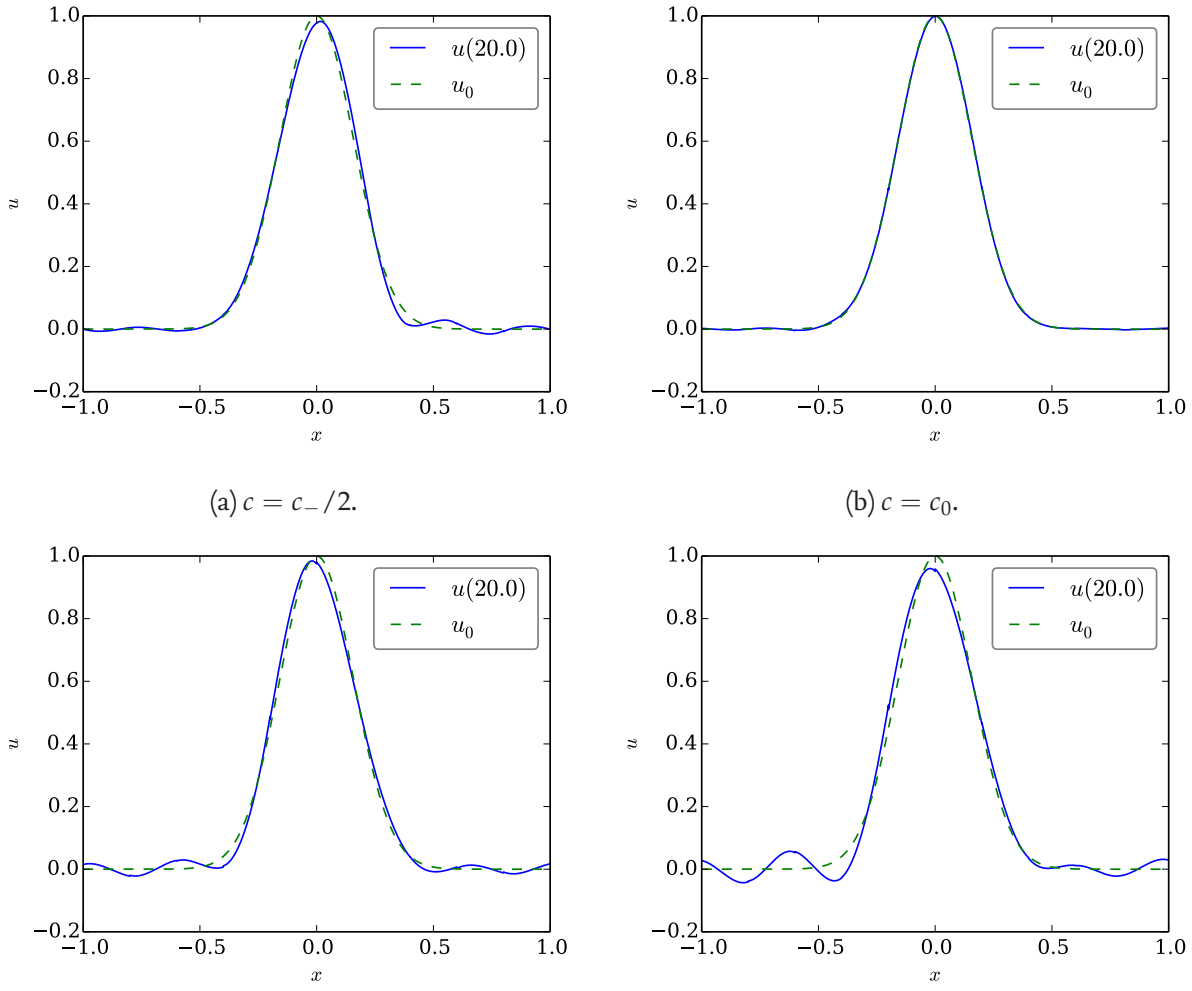
derived quantities like norms are computed exactly or approximately for these polynomials on each element. Therefore, Gauß-Legendre or Lobatto-Legendre quadrature rules are natural choices to compute L_2 norms. However, as shown in the previous sections, each choice of correction parameter for a CPR method is associated with a natural norm / scalar product, given by $\underline{\underline{M}} + \underline{\underline{K}}$. For $c = c_0 = 0$ and $c = c_{Hu}$, these scalar products are given by Gauß and Lobatto quadrature, respectively. Using a central flux, energy *in this specific norm* is conserved. By equivalence of norms in finite dimensional spaces, energy computed via other quadrature rules is bounded, but not necessarily conserved or non-increasing. This can be seen in Figures 3.2 and 3.3: The natural quadrature rules (Gauß for $c = c_0$ and Lobatto for $c = c_{Hu}$) yield exactly conserved energy, whereas other quadrature rules result in bounded oscillations of energy. Computing the norms $\|\cdot\|_{M+K}$ for $c = c_-/2$ and $c = c_{SD}$, the same conservation of energy is obtained, but not plotted here. However, as the solution $\underline{u}$ in each element represents a polynomial and not just some point values as in traditional FD methods, Gauß and Lobatto integration are standard choices to compute L_2 norms.

Using an upwind flux instead of a central flux, the corresponding results are shown in Figures 3.4, 3.5 and 3.6. As before, the results look like the ones of Vincent et al. [2011b]. The only difference is a stronger dissipation of energy for $c = c_-/2$ and $c = c_{Hu}$. This may be caused by the different time integrator and unknown values for the time steps used by Vincent et al. [2011b]. Again, the results obtained by using a Gauß-Legendre basis look very much the same.

As before, only the energy computed via the norm associated with the chosen correction is necessarily non-increasing. Due to dissipation by the upwind numerical flux, the corresponding energy decays. Other choices of quadrature rules still yield oscillating energy, decaying in the large.

A convergence study for a fixed number of elements $N = 10$ and varying polynomial degree p is plotted in Figure 3.7. The corresponding numerical values (rounded to two significant digits) are printed in Table 3.1. Both Gauß-Legendre and Lobatto-Legendre bases with an upwind numerical flux for different values of c are compared. In addition, the natural choice $\kappa = 0$ for each basis is considered, i.e. c_0 and c_{Hu} for Gauß and Lobatto quadrature, respectively. Nearly all parameters are the same as before, but the number of time steps is increased to $50,000$. For fixed c, the results for Gauß-Legendre and Lobatto-Legendre are similar, but for the natural choice $\kappa = 0$, Gauß-Legendre is clearly superior. All plots show clearly an approximately exponential decay of the error with increasing p up to about $p = 10$. There, higher precision than 64 bit floating point will probably lead to further decay.

Similar plots for a fixed polynomial degree $p = 4$ and varying number of elements N can be found in Figure 3.8 with corresponding numerical values (rounded to two significant digits) in Table 3.2. As before, for fixed c, the results are similar but for the natural choice $\kappa = 0$, the Gauß-Legendre basis is clearly superior. Note that both studies used the same total number of degrees of freedom and a high polynomial degree is superior compared to more number of elements for high precision. In this study, the limit error is not reached for any of the plotted number of elements.



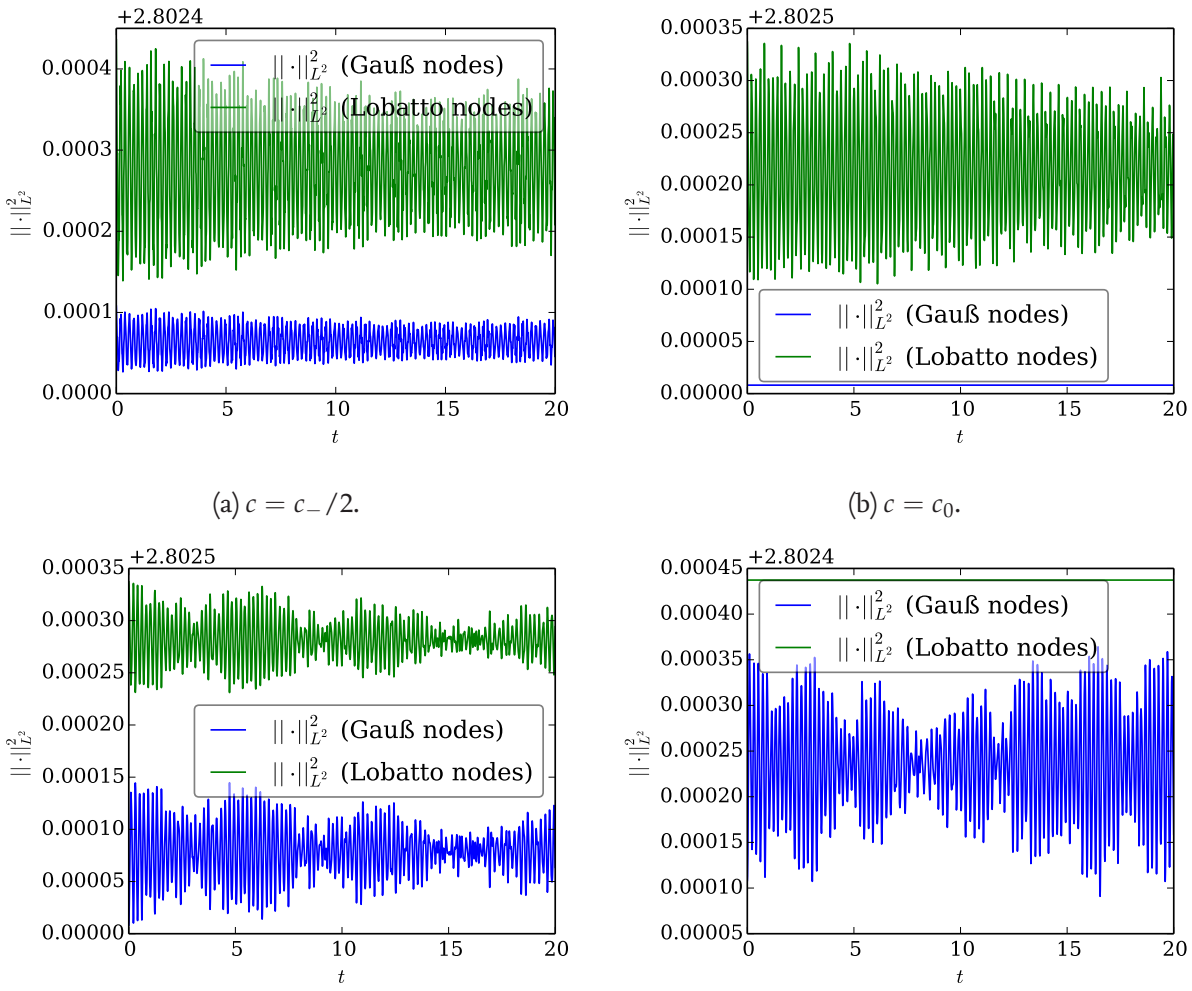
(a) $c = c_-/2$.

(b) $c = c_0$.

(c) $c = c_{SD}$.

(d) $c = c_{Hu}$.

Figure 3.1.: The numerical solution of constant velocity linear advection using SBP CPR methods with 10 elements, a Lobatto-Legendre basis of order $p = 3$ and a central numerical flux. Different values of c are used for the correction matrix $\underline{\underline{C}}$. The initial Gaussian profile u_0 is shown in green, the numerical solution is plotted in blue.



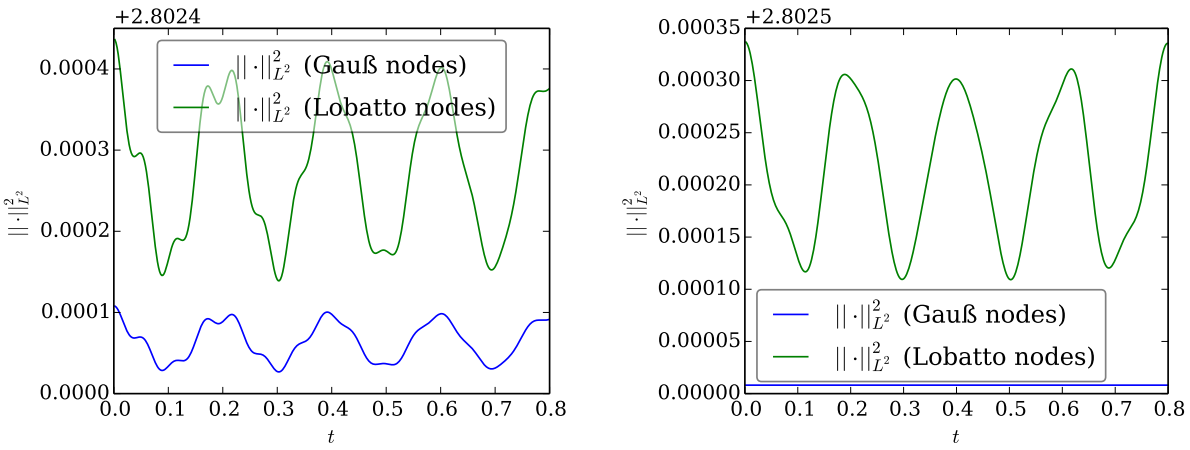
(a) $c = c_-/2$.

(b) $c = c_0$.

(c) $c = c_{SD}$.

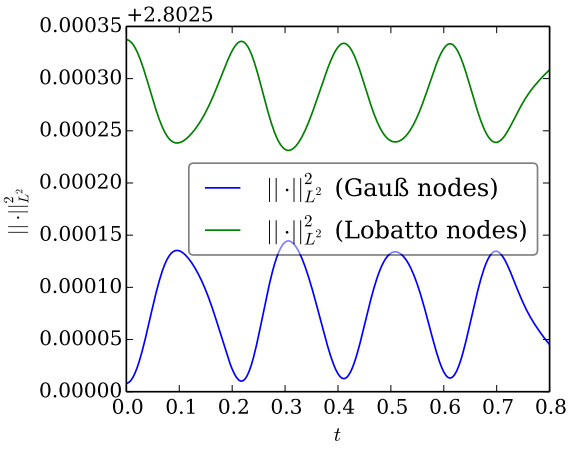
(d) $c = c_{Hu}$.

Figure 3.2.: Energy of numerical solution of constant velocity linear advection using SBP CPR methods with 10 elements, a Lobatto-Legendre basis of order $p = 3$ and a central numerical flux. Different values of c are used for the correction matrix $\underline{\underline{C}}$. The discrete energy $\|\underline{u}\|^2$ is computed using Gauß-Legendre (blue) and Lobatto-Legendre (green) quadrature with $p+1 = 4$ nodes in the full time interval $[0, 20]$.

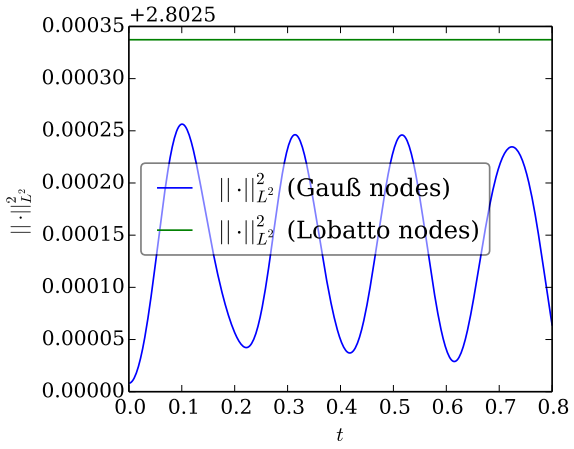


(a) $c = c_-/2$.

(b) $c = c_0$.

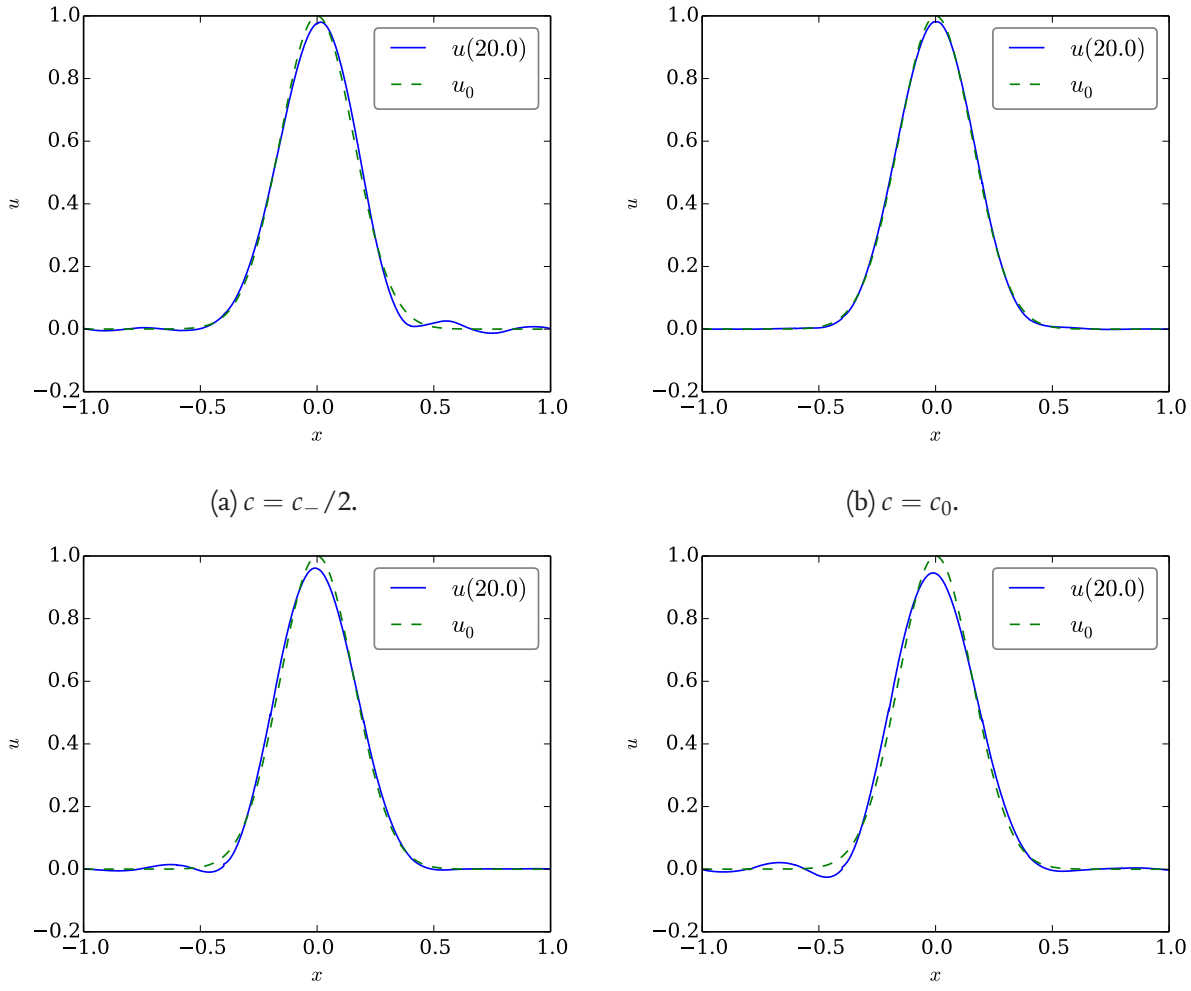


(c) $c = c_{SD}$.



(d) $c = c_{Hu}$.

Figure 3.3.: Energy of numerical solution of constant velocity linear advection using SBP CPR methods with 10 elements, a Lobatto-Legendre basis of order $p = 3$ and a central numerical flux. Different values of c are used for the correction matrix $\underline{\underline{C}}$. The discrete energy $\|\underline{u}\|^2$ is computed using Gauß-Legendre (blue) and Lobatto-Legendre (green) quadrature with $p + 1 = 4$ nodes in the time interval $[0, 0.8]$.



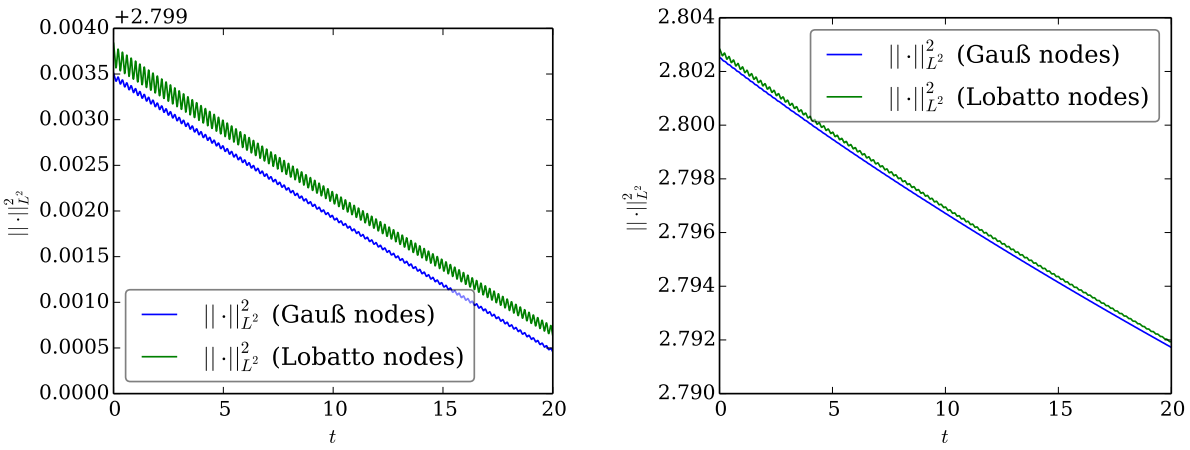
(a) $c = c_-/2$.

(b) $c = c_0$.

(c) $c = c_{SD}$.

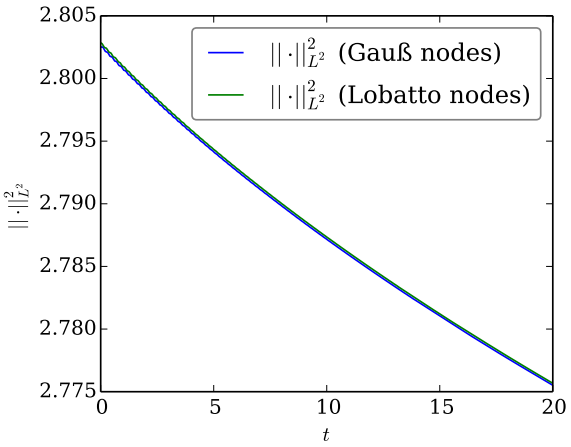
(d) $c = c_{Hu}$.

Figure 3.4.: The numerical solution of constant velocity linear advection using SBP CPR methods with 10 elements, a Lobatto-Legendre basis of order $p = 3$ and a upwind numerical flux. Different values of c are used for the correction matrix $\underline{\underline{C}}$. The initial Gaussian profile u_0 is shown in green, the numerical solution is plotted in blue.

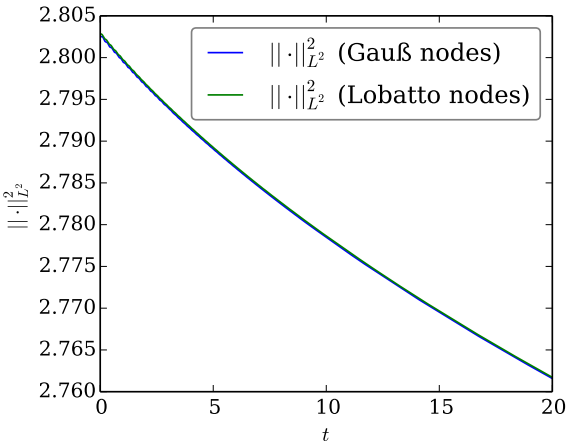


(a) $c = c_-/2$.

(b) $c = c_0$.

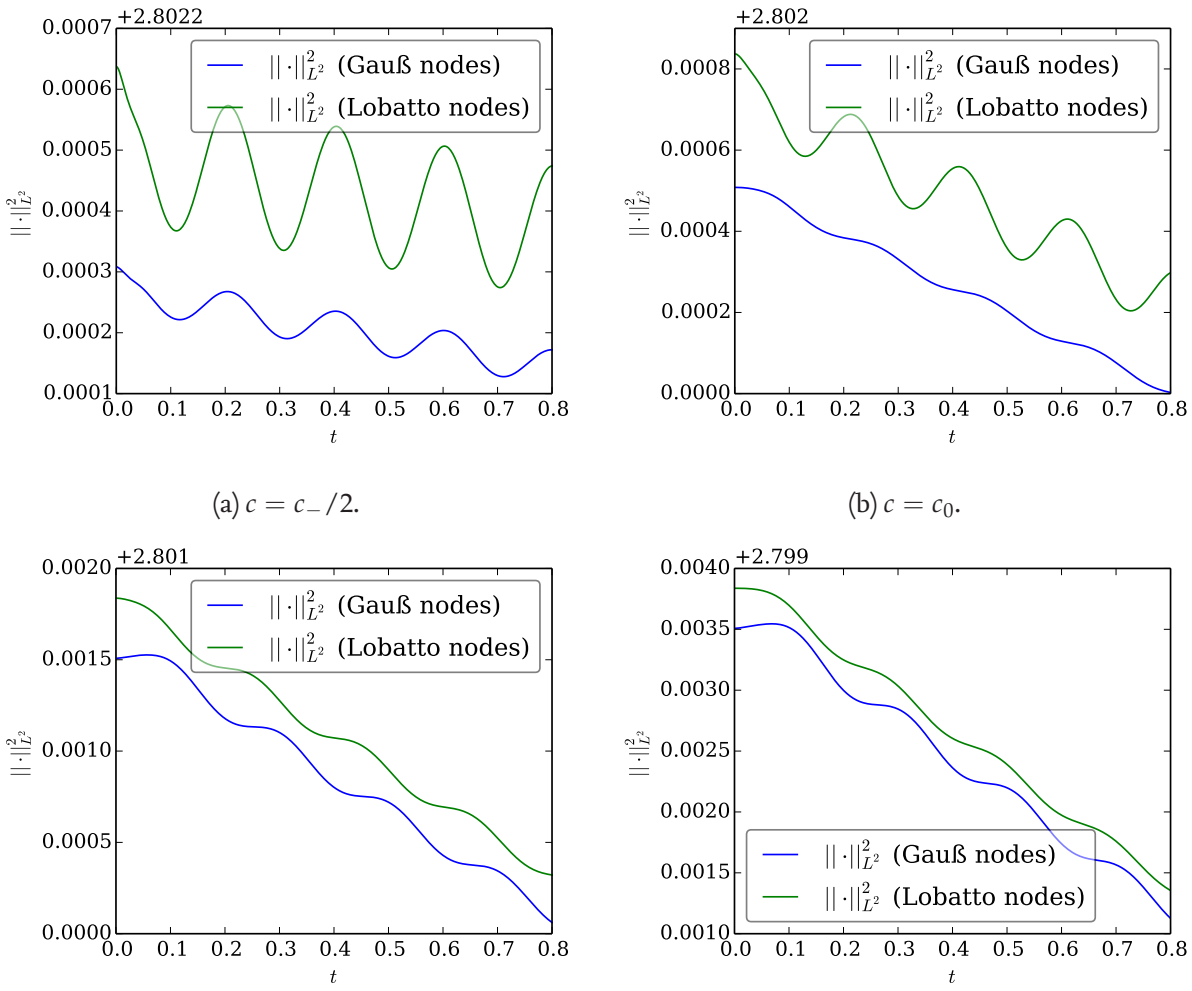


(c) $c = c_{SD}$.



(d) $c = c_{Hu}$.

Figure 3.5.: Energy of numerical solution of constant velocity linear advection using SBP CPR methods with 10 elements, a Lobatto-Legendre basis of order $p = 3$ and a upwind numerical flux. Different values of c are used for the correction matrix $\underline{\underline{C}}$. The discrete energy $\|\underline{u}\|^2$ is computed using Gauß-Legendre (blue) and Lobatto-Legendre (green) quadrature with $p + 1 = 4$ nodes in the full time interval $[0, 20]$.



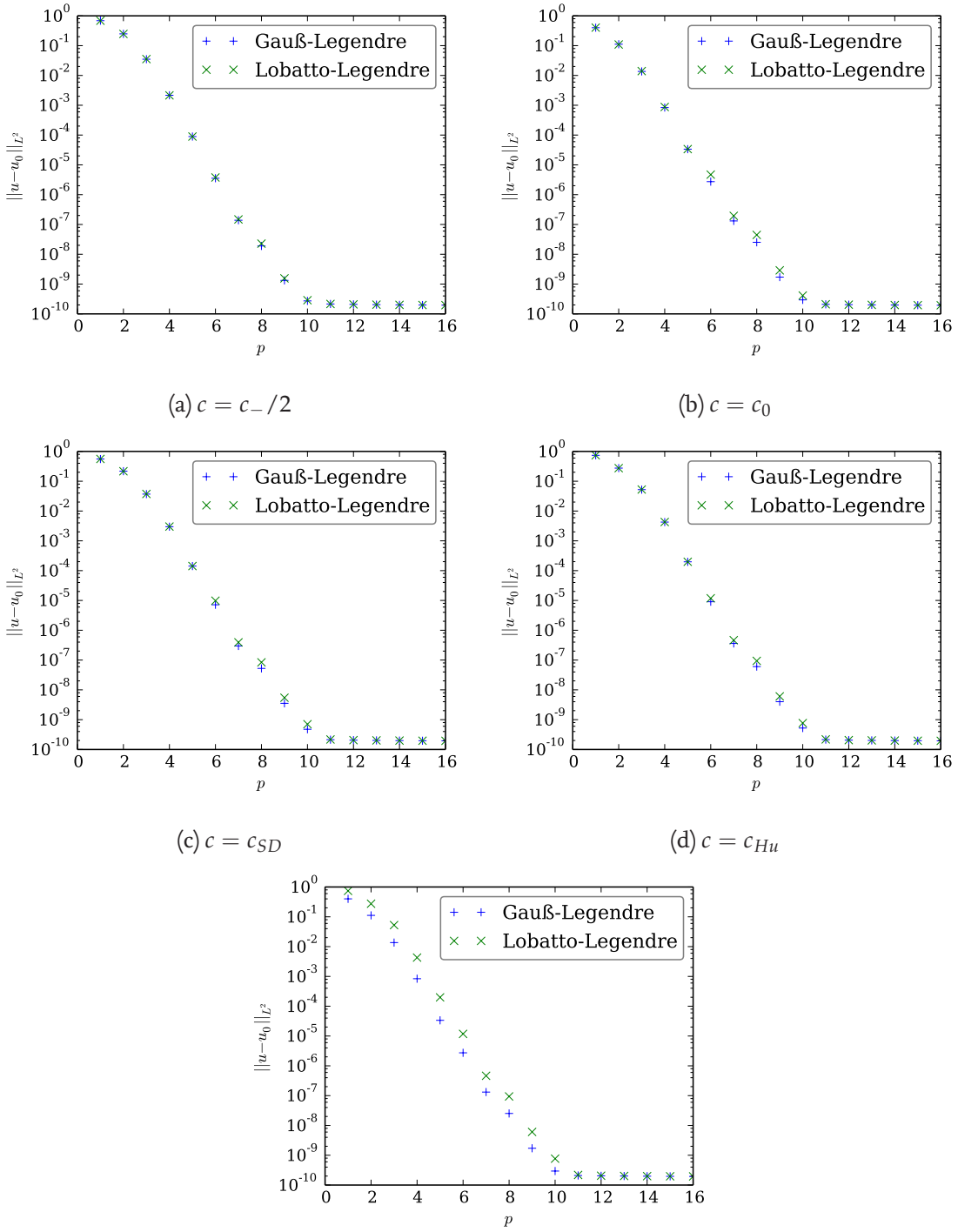
(a) $c = c_-/2$.

(b) $c = c_0$.

(c) $c = c_{SD}$.

(d) $c = c_{Hu}$.

Figure 3.6.: Energy of numerical solution of constant velocity linear advection using SBP CPR methods with 10 elements, a Lobatto-Legendre basis of order $p = 3$ and a upwind numerical flux. Different values of c are used for the correction matrix $\underline{\underline{C}}$. The discrete energy $\|\underline{u}\|^2$ is computed using Gauß-Legendre (blue) and Lobatto-Legendre (green) quadrature with $p + 1 = 4$ nodes in the time interval $[0, 0.8]$.



(a) $c = c_-/2$

(b) $c = c_0$

(c) $c = c_{SD}$

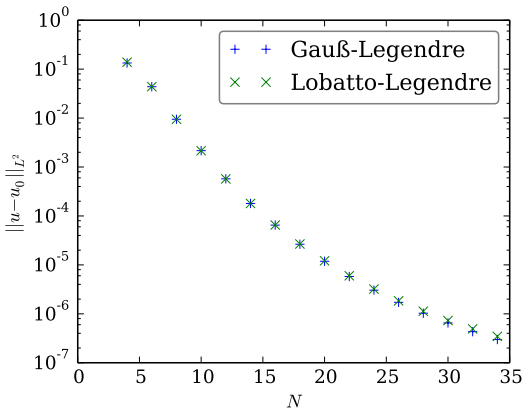
(d) $c = c_{Hu}$

(e) $\kappa = 0$

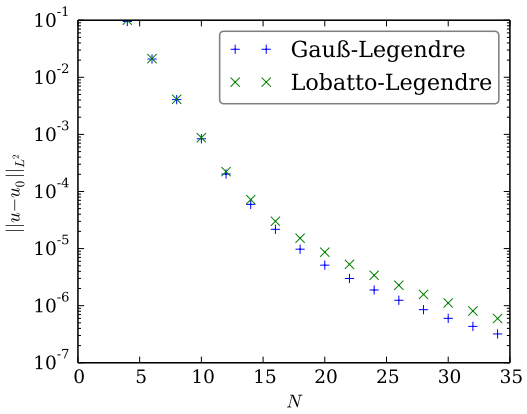
Figure 3.7.: L^2 errors of $u(20)$ for constant velocity linear advection using SBP CPR methods with $N = 10$ elements, a Gauß-Legendre and Lobatto-Legendre bases of varying order p and an upwind numerical flux. Different values of c are used for the correction matrix $\underline{\underline{C}}$.

Table 3.1.: L^2 errors of $u(20)$ for constant velocity linear advection using SBP CPR methods with $N = 10$ elements, Gauß-Legendre and Lobatto-Legendre bases of varying order p and an upwind numerical flux. Different values of c are used for the correction matrix $\underline{\underline{C}}$. Numerical values corresponding to Figure 3.7.

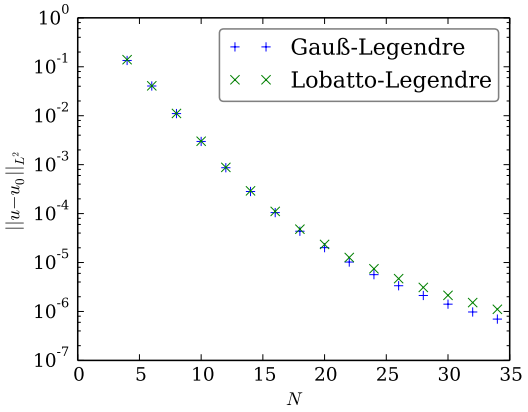
	$\|u - u_0\|_{L^2}$							
	$c = c_-/2$		$c = c_0$		$c = c_{SD}$		$c = c_{Hu}$	
p	Gauß	Lobatto	Gauß	Lobatto	Gauß	Lobatto	Gauß	Lobatto
1	6.88×10^{-1}	6.86×10^{-1}	4.00×10^{-1}	4.03×10^{-1}	5.55×10^{-1}	5.59×10^{-1}	7.40×10^{-1}	7.41×10^{-1}
2	2.49×10^{-1}	2.47×10^{-1}	1.12×10^{-1}	1.09×10^{-1}	2.17×10^{-1}	2.16×10^{-1}	2.75×10^{-1}	2.74×10^{-1}
3	3.51×10^{-2}	3.55×10^{-2}	1.36×10^{-2}	1.40×10^{-2}	3.70×10^{-2}	3.74×10^{-2}	5.22×10^{-2}	5.27×10^{-2}
4	2.16×10^{-3}	2.15×10^{-3}	8.38×10^{-4}	8.75×10^{-4}	2.98×10^{-3}	3.01×10^{-3}	4.26×10^{-3}	4.28×10^{-3}
5	8.92×10^{-5}	8.93×10^{-5}	3.36×10^{-5}	3.38×10^{-5}	1.42×10^{-4}	1.42×10^{-4}	1.98×10^{-4}	1.98×10^{-4}
6	3.56×10^{-6}	3.77×10^{-6}	2.73×10^{-6}	4.71×10^{-6}	7.16×10^{-6}	9.81×10^{-6}	9.04×10^{-6}	1.18×10^{-5}
7	1.39×10^{-7}	1.48×10^{-7}	1.32×10^{-7}	1.95×10^{-7}	2.97×10^{-7}	3.93×10^{-7}	3.61×10^{-7}	4.64×10^{-7}
8	1.90×10^{-8}	2.28×10^{-8}	2.52×10^{-8}	4.46×10^{-8}	5.25×10^{-8}	8.38×10^{-8}	6.03×10^{-8}	9.40×10^{-8}
9	1.33×10^{-9}	1.55×10^{-9}	1.71×10^{-9}	2.89×10^{-9}	3.53×10^{-9}	5.45×10^{-9}	4.00×10^{-9}	6.05×10^{-9}
10	2.73×10^{-10}	2.88×10^{-10}	2.95×10^{-10}	4.13×10^{-10}	4.77×10^{-10}	7.05×10^{-10}	5.23×10^{-10}	7.68×10^{-10}
11	2.15×10^{-10}	2.15×10^{-10}	2.09×10^{-10}	2.10×10^{-10}	2.13×10^{-10}	2.17×10^{-10}	2.14×10^{-10}	2.18×10^{-10}
12	2.09×10^{-10}	2.09×10^{-10}	2.04×10^{-10}	2.04×10^{-10}	2.05×10^{-10}	2.05×10^{-10}	2.05×10^{-10}	2.06×10^{-10}
13	2.04×10^{-10}	2.04×10^{-10}	2.01×10^{-10}	2.01×10^{-10}	2.01×10^{-10}	2.01×10^{-10}	2.02×10^{-10}	2.02×10^{-10}
14	2.01×10^{-10}	2.01×10^{-10}	1.98×10^{-10}	1.98×10^{-10}	1.99×10^{-10}	1.99×10^{-10}	1.99×10^{-10}	1.99×10^{-10}
15	1.98×10^{-10}	1.98×10^{-10}	1.96×10^{-10}	1.96×10^{-10}	1.97×10^{-10}	1.97×10^{-10}	1.97×10^{-10}	1.97×10^{-10}
16	1.96×10^{-10}	1.96×10^{-10}	1.95×10^{-10}	1.95×10^{-10}	1.95×10^{-10}	1.95×10^{-10}	1.95×10^{-10}	1.95×10^{-10}



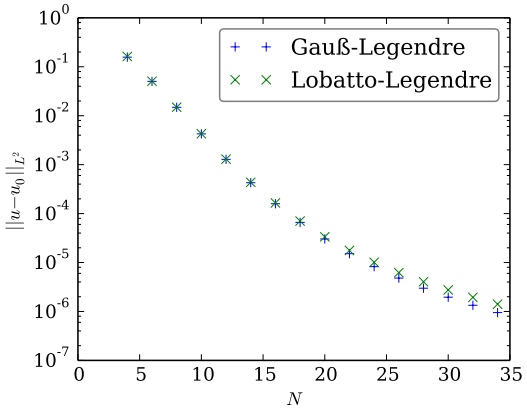
(a) $c = c_-/2$



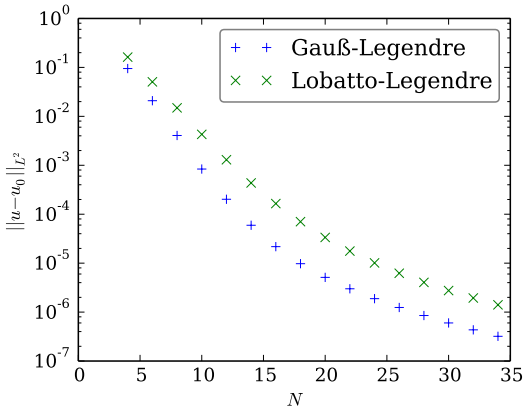
(b) $c = c_0$



(c) $c = c_{SD}$



(d) $c = c_{Hu}$



(e) $\kappa = 0$

Figure 3.8.: L^2 errors of $u(20)$ for constant velocity linear advection using SBP CPR methods with varying number of elements N, Gauß-Legendre and Lobatto-Legendre bases of order $p = 4$ and an upwind numerical flux. Different values of c are used for the correction matrix $\underline{\underline{C}}$.

Table 3.2.: L^2 errors of $u(20)$ for constant velocity linear advection using SBP CPR methods with varying number of elements N, Gauß-Legendre and Lobatto-Legendre bases of order $p = 4$ and an upwind numerical flux. Different values of c are used for the correction matrix $\underline{\underline{C}}$. Numerical values corresponding to Figure 3.8.

	$\|u - u_0\|_{L^2}$							
	$c = c_-/2$		$c = c_0$		$c = c_{SD}$		$c = c_{Hu}$	
p	Gauß	Lobatto	Gauß	Lobatto	Gauß	Lobatto	Gauß	Lobatto
4	1.34×10^{-1}	1.39×10^{-1}	9.48×10^{-2}	9.99×10^{-2}	1.35×10^{-1}	1.40×10^{-1}	1.57×10^{-1}	1.62×10^{-1}
6	4.37×10^{-2}	4.38×10^{-2}	2.09×10^{-2}	2.12×10^{-2}	4.05×10^{-2}	4.08×10^{-2}	5.04×10^{-2}	5.07×10^{-2}
8	9.46×10^{-3}	9.44×10^{-3}	4.06×10^{-3}	4.13×10^{-3}	1.11×10^{-2}	1.11×10^{-2}	1.49×10^{-2}	1.49×10^{-2}
10	2.16×10^{-3}	2.15×10^{-3}	8.38×10^{-4}	8.75×10^{-4}	2.98×10^{-3}	3.01×10^{-3}	4.26×10^{-3}	4.28×10^{-3}
12	5.73×10^{-4}	5.71×10^{-4}	2.02×10^{-4}	2.23×10^{-4}	8.65×10^{-4}	8.80×10^{-4}	1.28×10^{-3}	1.30×10^{-3}
14	1.80×10^{-4}	1.80×10^{-4}	5.95×10^{-5}	7.20×10^{-5}	2.83×10^{-4}	2.91×10^{-4}	4.27×10^{-4}	4.35×10^{-4}
16	6.50×10^{-5}	6.54×10^{-5}	2.18×10^{-5}	3.01×10^{-5}	1.05×10^{-4}	1.11×10^{-4}	1.59×10^{-4}	1.65×10^{-4}
18	2.64×10^{-5}	2.68×10^{-5}	9.75×10^{-6}	1.52×10^{-5}	4.36×10^{-5}	4.80×10^{-5}	6.60×10^{-5}	7.04×10^{-5}
20	1.19×10^{-5}	1.21×10^{-5}	5.12×10^{-6}	8.64×10^{-6}	2.01×10^{-5}	2.34×10^{-5}	3.03×10^{-5}	3.35×10^{-5}
22	5.79×10^{-6}	5.99×10^{-6}	3.00×10^{-6}	5.29×10^{-6}	1.02×10^{-5}	1.27×10^{-5}	1.52×10^{-5}	1.77×10^{-5}
24	3.05×10^{-6}	3.20×10^{-6}	1.88×10^{-6}	3.41×10^{-6}	5.65×10^{-6}	7.45×10^{-6}	8.24×10^{-6}	1.01×10^{-5}
26	1.72×10^{-6}	1.84×10^{-6}	1.24×10^{-6}	2.28×10^{-6}	3.36×10^{-6}	4.68×10^{-6}	4.81×10^{-6}	6.22×10^{-6}
28	1.03×10^{-6}	1.13×10^{-6}	8.51×10^{-7}	1.57×10^{-6}	2.12×10^{-6}	3.10×10^{-6}	2.99×10^{-6}	4.05×10^{-6}
30	6.51×10^{-7}	7.29×10^{-7}	6.00×10^{-7}	1.11×10^{-6}	1.41×10^{-6}	2.13×10^{-6}	1.96×10^{-6}	2.75×10^{-6}
32	4.32×10^{-7}	4.93×10^{-7}	4.34×10^{-7}	8.08×10^{-7}	9.76×10^{-7}	1.52×10^{-6}	1.34×10^{-6}	1.94×10^{-6}
34	2.98×10^{-7}	3.46×10^{-7}	3.20×10^{-7}	5.97×10^{-7}	6.98×10^{-7}	1.11×10^{-6}	9.49×10^{-7}	1.41×10^{-6}

3.9. Influence of time discretisation

Leaving the mathematical paradise of the previous sections and entering the real world of numerical methods, time discretisation plays an important role. For simplicity, an explicit Euler method for an SBP CPR semidiscretisation of the linear advection equation (3.13) is considered. Thus, one step of size Δt from $\underline{u}$ to $\underline{u}^+$ (in the standard element) can be written as

$$\underline{u}^+ = \underline{u} + \Delta t\, \partial_t \underline{u}. \tag{3.62}$$

Thus, the discrete norm after one time step is given by

$$\begin{aligned} \underline{u}^{+T} \underline{\underline{M}}\, \underline{u}^+ &= (\underline{u} + \Delta t\, \partial_t \underline{u})^T\, \underline{\underline{M}}\, (\underline{u} + \Delta t\, \partial_t \underline{u}) \\ &= \underline{u}^T \underline{\underline{M}}\, \underline{u} + 2\Delta t\, \underline{u}^T \underline{\underline{M}}\, \partial_t \underline{u} + \Delta t^2\, \partial_t \underline{u}^T \underline{\underline{M}}\, \partial_t \underline{u}. \end{aligned} \tag{3.63}$$

The computations leading to Lemma 3.2 result in an estimate of the second term $\underline{u}^T \underline{\underline{M}}\, \partial_t \underline{u} \leq 0$. Since $\underline{\underline{M}}$ is positive definite, the third term is non-negative (for $\Delta t > 0$). Thus, time discretisation introduces a growth of the discrete norm not considered in the previous calculations, possibly leading to stability issues. In order to get a decay of the discrete norm, the difference

$$\underline{u}^{+T} \underline{\underline{M}}\, \underline{u}^+ - \underline{u}^T \underline{\underline{M}}\, \underline{u} = 2\Delta t \left(\underline{u} + \frac{\Delta t}{2} \partial_t \underline{u} \right)^T \underline{\underline{M}}\, \partial_t \underline{u} \tag{3.64}$$

must be estimated. Inserting the SBP CPR semidiscretisation for the linear advection equation with constant velocity yields

$$\begin{aligned} &\left(\underline{u} + \frac{\Delta t}{2} \partial_t \underline{u} \right)^T \underline{\underline{M}}\, \partial_t \underline{u} = \left(\underline{u} - \frac{\Delta t}{2} \underline{\underline{D}}\, \underline{u} - \frac{\Delta t}{2} \underline{\underline{C}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) \right)^T \underline{\underline{M}}\, \left(-\underline{\underline{D}}\, \underline{u} - \underline{\underline{C}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) \right) \\ &= -\underline{u}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} - \underline{u}^T \underline{\underline{M}}\, \underline{\underline{C}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) \\ &\quad + \frac{\Delta t}{2} \left[\underline{u}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} + (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u})^T \underline{\underline{C}}^T \underline{\underline{M}}\, \underline{\underline{C}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) + 2\underline{u}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{C}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) \right]. \end{aligned} \tag{3.65}$$

Inserting $\underline{\underline{C}} = \underline{\underline{M}}^{-1} \underline{\underline{R}}^T \underline{\underline{B}}$, the right hand side becomes

$$\begin{aligned} &- \underline{u}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} - \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) \\ &+ \frac{\Delta t}{2} \left[\underline{u}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} + (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u})^T \underline{\underline{B}}^T \underline{\underline{R}}\, \underline{\underline{M}}^{-1} \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) + 2\underline{u}^T \underline{\underline{D}}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) \right]. \end{aligned} \tag{3.66}$$

Using the SBP property $\underline{\underline{M}}\, \underline{\underline{D}} + \underline{\underline{D}}^T \underline{\underline{M}} = \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}$ results in

$$\begin{aligned} &\underline{u}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{u} - \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}\, \underline{u} - \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) \\ &+ \frac{\Delta t}{2} \left[\underline{u}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} + (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u})^T \underline{\underline{B}}^T \underline{\underline{R}}\, \underline{\underline{M}}^{-1} \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) + 2\underline{u}^T \underline{\underline{D}}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) \right]. \end{aligned} \tag{3.67}$$

Thus, adding these expressions, the left hand side LHS can be expressed as

$$\begin{aligned} &2\, LHS \\ &= \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}\, \underline{u} - 2\underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{f}^{\text{num}} \\ &\quad + \Delta t \left[\underline{u}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} + (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u})^T \underline{\underline{B}}^T \underline{\underline{R}}\, \underline{\underline{M}}^{-1} \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) + 2\underline{u}^T \underline{\underline{D}}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{f}^{\text{num}} - \underline{\underline{R}}\, \underline{u}) \right]. \end{aligned} \tag{3.68}$$

If all summands on the right hand side would involve only boundary terms, an estimate similar to those in previous sections would be possible. Unfortunately, the volume term $\underline{u}^T \underline{\underline{D}}^T \underline{\underline{M}} \underline{\underline{D}} \underline{u}$ does not seem to allow such an estimate. Therefore, an estimate leading to fully discrete stability for an explicit Euler method does not seem to be possible in this straightforward calculation. Thus, for practical calculations, a time discretisation with high accuracy should be chosen in order to avoid stability issues. Note that the non-positive stability result for the explicit Euler methods extends directly to standard SSP time discretisations consisting of convex combinations of Euler steps.

4 Nonlinear stability for Burgers' equation

Stability properties for linear and nonlinear problems can be very different. Thus, although stability for linear advection with constant velocity can be proven for several SBP CPR schemes (see Theorem 3.5), these results do not imply nonlinear stability. For simplicity, the (inviscid) Burgers' equation

$$\partial_t u + \partial_x \frac{u^2}{2} = 0 \tag{4.1}$$

in one space dimension with periodic boundary conditions and appropriate initial condition is considered.

This chapter has been published by order of Professor Sonar [Ranocha et al., 2015b, 2016].

4.1. Nonlinear stability

A straightforward application of an SBP CPR method can be written as

$$\partial_t \underline{u} + \frac{1}{2}\underline{\underline{D}}\,\underline{u^2} + \underline{\underline{C}}\,(\underline{f}^{\text{num}} - \frac{1}{2}\underline{\underline{R}}\,\underline{u^2}) = 0 \tag{4.2}$$

for the standard element. Estimating the discrete norm similar to the previous sections results in

$$\underline{u}^T \underline{\underline{M}}\,\partial_t \underline{u} = -\frac{1}{2}\underline{u}^T \underline{\underline{M}}\,\underline{\underline{D}}\,\underline{u^2} - \underline{u}^T \underline{\underline{M}}\,\underline{\underline{C}}\,(\underline{f}^{\text{num}} - \frac{1}{2}\underline{\underline{R}}\,\underline{u^2}). \tag{4.3}$$

Applying the SBP property and $\underline{\underline{M}}\,\underline{\underline{C}} = \underline{\underline{R}}^T \underline{\underline{B}}$ yields

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = \frac{1}{2}\underline{u}^T \underline{\underline{D}}^T \underline{\underline{M}}\,\underline{u^2} - \frac{1}{2}\underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u^2} - \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\,(\underline{f}^{\text{num}} - \frac{1}{2}\underline{\underline{R}}\,\underline{u^2}). \tag{4.4}$$

Unfortunately, the nonlinear flux does not allow a cancellation of boundary terms as in the linear case. A possibility to overcome this problem in the setting of DG spectral element methods was proposed by Gassner [2013]. There, he uses Lobatto-Legendre interpolation polynomials as nodal basis and a skew-symmetric form of the conservation law

$$\partial_t u + \alpha \partial_x \frac{u^2}{2} + (1-\alpha) u \partial_x u = 0, \quad 0 \leq \alpha \leq 1. \tag{4.5}$$

Thus, the divergence of the flux is written as a convex combination of the two terms $\partial_x \frac{u^2}{2}$ and $u\partial_x u$ which are exactly equal if the product rule of differentiation is valid for ∂_x. The discrete derivative operator $\underline{\underline{D}}$ does not fulfil this product rule and therefore, the split operator form (4.5) can be regarded as the standard conservative form (4.1) with an additional correction term

$$\partial_t u + \partial_x \frac{u^2}{2} + (1-\alpha)\left(u\partial_x u - \partial_x \frac{u^2}{2}\right) = 0. \tag{4.6}$$

Using the SBP CPR semidiscretisation for this equation yields

$$\partial_t \underline{u} = -\frac{1}{2}\underline{\underline{D}}\,\underline{u^2} - (1-\alpha)(\underline{u}\,\underline{\underline{D}}\,\underline{u} - \frac{1}{2}\underline{\underline{D}}\,\underline{u^2}) - \underline{\underline{C}}\,(\underline{f}^{\text{num}} - \frac{1}{2}\underline{\underline{R}}\,\underline{u^2}) = 0. \tag{4.7}$$

Here, $\underline{\underline{u}} = \operatorname{diag}(\underline{u})$ denotes the matrix representing multiplication with u. Now, multiplication with $\underline{u}^T\underline{\underline{M}}$ results in

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = -\frac{\alpha}{2}\underline{u}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{u^2} - (1-\alpha)\underline{u}^T\underline{\underline{M}}\,\underline{\underline{u}}\,\underline{\underline{D}}\,\underline{u} - \underline{u}^T\underline{\underline{M}}\,\underline{\underline{C}}\,(\underline{f}^{\mathrm{num}} - \frac{1}{2}\underline{\underline{R}}\,\underline{u^2}). \tag{4.8}$$

Using $\underline{\underline{C}} = \underline{\underline{M}}^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$ and $\underline{u^2} = \underline{\underline{u}}\,\underline{u}$, this can be written as

$$\frac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = -\frac{\alpha}{2}\underline{u}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{\underline{u}}\,\underline{u} - (1-\alpha)\underline{u}^T\underline{\underline{M}}\,\underline{\underline{u}}\,\underline{\underline{D}}\,\underline{u} - \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\mathrm{num}} + \frac{1}{2}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{\underline{u}}\,\underline{u}. \tag{4.9}$$

Since a nodal Lobatto-Legendre basis with lumped mass matrix is chosen, both $\underline{\underline{u}}$ and $\underline{\underline{M}}$ are diagonal and therefore commute

$$\frac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = -\frac{\alpha}{2}\underline{u}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{\underline{u}}\,\underline{u} - (1-\alpha)\underline{u}^T\underline{\underline{u}}\,\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{u} - \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\mathrm{num}} + \frac{1}{2}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{\underline{u}}\,\underline{u}. \tag{4.10}$$

Application of the SBP property results in

$$\begin{aligned}\frac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 =& \frac{\alpha}{2}\underline{u}^T\underline{\underline{D}}^T\underline{\underline{M}}\,\underline{\underline{u}}\,\underline{u} - \frac{\alpha}{2}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{\underline{u}}\,\underline{u} - (1-\alpha)\underline{u}^T\underline{\underline{u}}\,\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{u}\\ &- \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\mathrm{num}} + \frac{1}{2}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{\underline{u}}\,\underline{u}.\end{aligned} \tag{4.11}$$

Choosing $\alpha = \frac{2}{3}$, $\frac{\alpha}{2} = 1-\alpha$ and the volume terms cancel out

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = \frac{1}{6}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{\underline{u}}\,\underline{u} - \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\mathrm{num}}. \tag{4.12}$$

Thus, Gassner [2013] is able to estimate the rate of change in one element as

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = \frac{1}{6}(u_R^3 - u_L^3) - (u_R f_R^{\mathrm{num}} - u_L f_L^{\mathrm{num}}), \tag{4.13}$$

where the indices L and R denote the values at the left and right boundary points, respectively. Lobatto-Legendre quadrature is necessary for this calculation, because the boundary points are nodes for the basis and therefore the restriction of u^2 to the boundary is the square of the restriction of u to the boundary, i.e. $\underline{\underline{R}}\,\underline{u^2} = (\underline{\underline{R}}\,\underline{u})^2$. In general, this is false for other nodes, as for example Gauß-Legendre quadrature.

Continuing the investigation, using periodic boundary conditions and summing over all elements, the contribution of one boundary can be expressed as

$$\frac{1}{6}(u_-^3 - u_+^3) - (u_- - u_+)f^{\mathrm{num}}, \tag{4.14}$$

where the indices $-$ and $+$ indicate the values from the left and right element, respectively. With the choice of

$$f^{\mathrm{num}} = \frac{1}{2}\left(\frac{u_+^2}{2} + \frac{u_-^2}{2}\right) - \lambda(u_+ - u_-) \tag{4.15}$$

as numerical flux, one can estimate this contribution like Gassner [2013]

$$\begin{aligned}&\frac{1}{6}(u_-^3 - u_+^3) + \frac{1}{4}(u_+ - u_-)(u_+^2 + u_-^2) - \lambda(u_+ - u_-)^2\\ &= \frac{1}{6}(u_-^3 - u_+^3) + \frac{1}{4}(u_+^3 - u_+^2u_- + u_+u_-^2 - u_-^3) - \lambda(u_+ - u_-)^2\\ &= \frac{1}{12}(u_+^3 - 3u_+^2u_- + 3u_+u_-^2 - u_-^3) - \lambda(u_+ - u_-)^2\\ &= (u_+ - u_-)^2\left(\frac{u_+ - u_-}{12} - \lambda\right).\end{aligned} \tag{4.16}$$

Thus, $\lambda \geq \frac{u_+ - u_-}{12}$ ensures $\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 \leq 0$, and therefore stability. This proves the following

Lemma 4.1 (see also Gassner [2013]). *If the numerical flux f^{num} satisfies*

$$\frac{1}{6}(u_-^3 - u_+^3) - (u_- - u_+) f^{\text{num}}(u_-, u_+) \leq 0, \tag{4.17}$$

the CPR method with nodal Lobatto-Legendre basis and $\underline{\underline{C}} = \underline{\underline{M}}^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$ for the skew-symmetric inviscid Burgers' equation (4.5) with correction parameter $\alpha = \frac{2}{3}$, written as

$$\partial_t \underline{u} + \underline{\underline{D}}\frac{1}{2}\underline{u^2} + \frac{1}{3}\left(\underline{u}\,\underline{\underline{D}}\,\underline{u} - \underline{\underline{D}}\frac{1}{2}\underline{u^2}\right) + \underline{\underline{C}}\left(\underline{f}^{\text{num}} - \underline{\underline{R}}\frac{1}{2}\underline{u^2}\right) = 0, \tag{4.18}$$

is stable in the discrete norm $\|\cdot\|_M$ induced by $\underline{\underline{M}}$.

As remarked above, this stability result is based on Lobatto-Legendre nodes including the boundaries. To get stability for a general SBP basis, further corrections are necessary. To the author's knowledge, this is a new idea and not published anywhere else. Recalling equation (4.12), the contribution of one boundary is

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = \frac{1}{6}(u_-(u^2)_- - u_+(u^2)_+) + (u_+ - u_-) f^{\text{num}}. \tag{4.19}$$

In general, multiplication and restriction to the boundary do not commute, i.e. $(u_R)^2 \neq (u^2)_R$, as mentioned above. In comparison with the estimate (4.13) of Gassner [2013],

$$\frac{1}{6}(u_-(u^2)_- - u_+(u^2)_+) - \frac{1}{6}(u_-^3 - u_+^3) \tag{4.20}$$

appears as additional term on the right hand side, possibly leading to instability. Therefore, an SBP CPR method with corrected divergence (skew-symmetric form) and corrected boundary terms is proposed

$$\partial_t \underline{u} + \frac{\alpha}{2}\underline{\underline{D}}\,\underline{u^2} + (1-\alpha)\underline{u}\,\underline{\underline{D}}\,\underline{u} + \underline{\underline{C}}\left(\underline{f}^{\text{num}} - \frac{\beta}{2}\underline{\underline{R}}\,\underline{u^2} - \frac{1-\beta}{2}(\underline{\underline{R}}\,\underline{u})^2\right) = 0, \quad 0 \leq \alpha, \beta \leq 1. \tag{4.21}$$

Setting $\alpha = \frac{2}{3}$ and repeating the calculations as above results in

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = -\underline{u}^T\underline{\underline{R}}\,\underline{\underline{B}}\,\underline{f}^{\text{num}} - \left(\frac{1}{3} - \frac{\beta}{2}\right)\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u^2} + \frac{1-\beta}{2}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u})^2. \tag{4.22}$$

Therefore, setting $\beta = \frac{2}{3}$ results in

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 = -\underline{u}^T\underline{\underline{R}}\,\underline{\underline{B}}\,\underline{f}^{\text{num}} + \frac{1}{6}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u})^2, \tag{4.23}$$

and the contribution of one boundary is the same as for Lobatto-Legendre nodes in (4.14). This proves the following

Lemma 4.2. *If the numerical flux f^{num} satisfies*

$$\frac{1}{6}(u_-^3 - u_+^3) - (u_- - u_+) f^{\text{num}}(u_-, u_+) \leq 0, \tag{4.24}$$

the SBP CPR method with $\underline{\underline{C}} = \underline{\underline{M}}^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$ *for the inviscid Burgers' equation* (4.1) *with correction terms for the divergence and restriction to the boundary, written as*

$$\partial_t \underline{u} + \underline{\underline{D}}\frac{1}{2}\underline{u^2} + \frac{1}{3}\left(\underline{u}\,\underline{\underline{D}}\,\underline{u} - \underline{\underline{D}}\frac{1}{2}\underline{u^2}\right) + \underline{\underline{C}}\left(\underline{f}^{\text{num}} - \underline{\underline{R}}\frac{1}{2}\underline{u^2} - \frac{1}{3}\left(\frac{1}{2}(\underline{\underline{R}}\,\underline{u})^2 - \frac{1}{2}\underline{\underline{R}}\,\underline{u^2}\right)\right) = 0, \quad (4.25)$$

is stable in the discrete norm $\|\cdot\|_M$ *induced by* $\underline{\underline{M}}$.

The motivation to introduce the skew-symmetric form (the divergence correction) as described by Gassner [2013] (see also inter alia Fisher et al. [2013], Svärd and Nordström [2014], Fernández et al. [2014b]) was the invalid product rule for the discrete derivative operator $\underline{\underline{D}}$. In view of the previous Lemma, the inexactness of *discrete multiplication* is stressed, resulting in both an invalid product rule for polynomial bases and incorrect restriction to the boundary for nodal bases not including boundary points.

4.2. Conservation

In order to be useful, the semidiscretisation (4.25) also has to be conservative. As in Lemma 3.1, multiplication with the constant function, represented as $\underline{1}$, yields

$$\underline{1}^T\underline{\underline{M}}\,\partial_t\underline{u} = -\underline{1}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{f} - \underline{1}^T\underline{\underline{M}}\,\underline{c}_{div} - \underline{1}^T\underline{\underline{M}}\,\underline{\underline{C}}\,(\underline{f}^{\text{num}} - \underline{\underline{R}}\,\underline{f} - \underline{c}_{res}). \quad (4.26)$$

Here and in the following, $\underline{c}_{div}$ and $\underline{c}_{res}$ denote correction terms for the divergence and restriction, respectively. Using $\underline{1}^T\underline{\underline{M}}\,\underline{\underline{C}} = \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}$ and the SBP property results in

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = \underline{1}^T\underline{\underline{D}}^T\underline{\underline{M}}\,\underline{f} - \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{f} - \underline{1}^T\underline{\underline{M}}\,\underline{c}_{div} - \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{f}^{\text{num}} - \underline{\underline{R}}\,\underline{f} - \underline{c}_{res}). \quad (4.27)$$

Since the discrete derivative is exact for constant functions, $\underline{\underline{D}}\,\underline{1} = 0$, and the rate of change can be expressed as

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = -\underline{1}^T\underline{\underline{M}}\,\underline{c}_{div} - \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\text{num}} + \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{c}_{res}. \quad (4.28)$$

Inserting the correction terms

$$\underline{c}_{div} = \frac{1}{3}\left(\underline{u}\,\underline{\underline{D}}\,\underline{u} - \frac{1}{2}\underline{\underline{D}}\,\underline{u}\,\underline{u}\right), \qquad \underline{c}_{res} = \frac{1}{6}\left((\underline{\underline{R}}\,\underline{u})^2 - \underline{\underline{R}}\,\underline{u}\,\underline{u}\right), \quad (4.29)$$

$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u}$ is rewritten as

$$-\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\text{num}} - \frac{1}{3}\underline{1}^T\underline{\underline{M}}\,\underline{u}\,\underline{\underline{D}}\,\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{u}\,\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u})^2 - \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u}\,\underline{u}. \quad (4.30)$$

For diagonal-norm SBP operators, both $\underline{\underline{M}}$ and $\underline{u}$ are diagonal and therefore commute. Using $\underline{u}\,\underline{1} = \underline{u}$ results in

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = -\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\text{num}} - \frac{1}{3}\underline{u}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{u}\,\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u})^2 - \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u}\,\underline{u}. \quad (4.31)$$

The SBP property yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = &-\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\text{num}} + \frac{1}{3}\underline{u}^T\underline{\underline{D}}^T\underline{\underline{M}}\,\underline{u} - \frac{1}{3}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u} \\ &+ \frac{1}{6}\underline{1}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{u}\,\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u})^2 - \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u}\,\underline{u}.\end{aligned} \quad (4.32)$$

Adding the last two equations and multiplying by one half results in

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = -\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\text{num}} - \frac{1}{6}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{u}\,\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u})^2 - \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u}\,\underline{u}. \tag{4.33}$$

Again, by using the SBP property

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = &-\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\text{num}} - \frac{1}{6}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u}\\ &- \frac{1}{6}\underline{1}^T\underline{\underline{D}}^T\underline{\underline{M}}\,\underline{u}\,\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u}\,\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u})^2 - \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u}\,\underline{u}.\end{aligned} \tag{4.34}$$

Gathering terms and using $\underline{\underline{D}}\,\underline{1} = 0$, this can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = -\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\text{num}} - \frac{1}{6}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u})^2. \tag{4.35}$$

Finally, since

$$\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{u} = u_R \cdot u_R - u_L \cdot u_L = 1 \cdot u_R^2 - 1 \cdot u_L^2 = \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,(\underline{\underline{R}}\,\underline{u})^2, \tag{4.36}$$

this reduces to the same equation as in the proof of Lemma 3.1. Thus, the following Lemma is proved

Lemma 4.3. *If $\underline{1}^T\underline{\underline{M}}\,\underline{\underline{C}} = \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}$, the SBP CPR method (4.25) for the inviscid Burgers' equation (4.1) is conservative.*

As Lemma 3.1, Lemma 4.3 proofs conservation across elements. On a sub-element level, conservation for diagonal-norm SBP operators (including boundary nodes) and conservation laws in split form has been proven by Fisher et al. [2013] in the context of the Lax-Wendroff theorem.

4.3. Numerical fluxes

In the following, some numerical fluxes for Burgers' equation are investigated. Gassner [2013] considered fluxes of the form (4.15). Choosing $\lambda = (u_+ - u_-)/12$ leads to the *energy conservative* (ECON) flux of Gassner [2013]

$$f^{\text{num}}(u_-, u_+) = \frac{1}{4}(u_+^2 + u_-^2) - \frac{(u_+ - u_-)^2}{12}. \tag{4.37}$$

With this choice, the contribution of the boundary terms vanishes and therefore the energy $\|u\|^2$ remains constant. Since the energy is also an entropy, this will result in unphysical solutions after the formation of shocks.

The choice $\lambda = |u_+ + u_-|/2$ results in *Roe's* flux

$$f^{\text{num}}(u_-, u_+) = \frac{1}{4}(u_+^2 + u_-^2) - \frac{|u_+ - u_-|}{2}(u_+ - u_-). \tag{4.38}$$

Unfortunately, the contribution (4.16) is not guaranteed to be non-negative, since $|u_+ + u_-| \geq (u_+ - u_-)/6$ is possible, e.g. for $u_+ = -u_- > 0$. Therefore, this choice does not imply stability.

Finally, Gassner [2013] considered the *local Lax-Friedrichs* (LLF) flux with parameter $\lambda = \frac{\max(|u_+|,|u_-|)}{2}$

$$f^{\text{num}}(u_-, u_+) = \frac{1}{4}(u_+^2 + u_-^2) - \frac{\max(|u_+|, |u_-|)}{2}(u_+ - u_-), \tag{4.39}$$

leading to entropy stability, since $\max(|u_+|, |u_-|) \geq |u_+| + |u_-| \geq (u_+ - u_-)/6$.

Another possible numerical flux is *Osher's* flux (see Toro [2009, section 12.1.4])

$$f^{\text{num}}(u_-, u_+) = \begin{cases} \frac{u_-^2}{2}, & u_+, u_- > 0, \\ \frac{u_+^2}{2}, & u_+, u_- < 0, \\ \frac{u_+^2}{2} + \frac{u_-^2}{2}, & u_- \geq 0 \geq u_+, \\ 0, & u_- \leq 0 \leq u_+. \end{cases} \tag{4.40}$$

Inserting this flux in the condition of Lemma 4.2 for the case $u_+, u_- > 0$ leads to

$$\begin{aligned} &\frac{1}{6}(u_-^3 - u_+^3) - (u_- - u_+)\frac{u_-^2}{2} = -\frac{1}{3}u_-^3 - \frac{1}{6}u_+^3 + \frac{1}{2}u_+u_-^2 \\ &\leq -\frac{1}{3}u_-^3 - \frac{1}{6}u_+^3 + \frac{1}{2}\left(\frac{1}{3}u_+^3 + \frac{2}{3}u_-^3\right) = 0, \end{aligned} \tag{4.41}$$

where Young's inequality

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}, \quad a, b > 0, \quad \frac{1}{p} + \frac{1}{q} = 1 \tag{4.42}$$

was used. The case $u_+, u_- < 0$ is similar. If $u_- \geq 0 \geq u_+$, the condition of Lemma 4.2 reads

$$\frac{1}{6}(u_-^3 - u_+^3) - \frac{1}{2}(u_- - u_+)(u_+^2 + u_-^2) = -\frac{1}{3}u_-^3 + \frac{1}{3}u_+^3 - \frac{1}{2}u_+^2u_- + \frac{1}{2}u_+u_-^2 \leq 0, \tag{4.43}$$

since each term is not positive. Finally, for $u_- \leq 0 \leq u_+$, the contribution $(u_-^3 - u_+^3)/6$ is again not positive. Thus, this numerical flux results in a stable scheme.

4.4. Summary and numerical results

The results are summed up in the following

Theorem 4.4. *If the numerical flux f^{num} satisfies*

$$\frac{1}{6}(u_-^3 - u_+^3) - (u_- - u_+)f^{\text{num}}(u_-, u_+) \leq 0, \tag{4.44}$$

then an SBP CPR method with $\underline{\underline{C}} = \underline{\underline{M}}^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$ and correction terms for both divergence and restriction to the boundary

$$\partial_t \underline{u} + \underline{\underline{D}}\,\underline{\frac{1}{2}u^2} + \frac{1}{3}\left(\underline{u}\,\underline{\underline{D}}\,\underline{u} - \underline{\underline{D}}\,\underline{\frac{1}{2}u^2}\right) + \underline{\underline{C}}\left(\underline{f}^{\text{num}} - \underline{\underline{R}}\,\underline{\frac{1}{2}u^2} - \frac{1}{3}\left(\frac{1}{2}(\underline{\underline{R}}\,\underline{u})^2 - \frac{1}{2}\underline{\underline{R}}\,\underline{u^2}\right)\right) = 0, \tag{(4.25)}$$

for the inviscid Burgers' equation (4.1) is both conservative and stable in the discrete norm $\|\cdot\|_M$ induced by $\underline{\underline{M}}$. Numerical fluxes fulfilling this condition are inter alia

- *the energy conservative (ECON) flux (4.37),*
- *the local Lax-Friedrichs (LLF) flux (4.39),*
- *and Osher's flux (4.40).*

Of course, the ECON flux should not be used in situations involving discontinuities, as shown in the following numerical examples. The setting is the same as in the case considered by Gassner [2013], i.e. the inviscid Burgers' equation (4.1) in the domain $[0, 2]$ with periodic boundary conditions is solved. The initial condition is

$$u(0, x) = u_0(x) = \sin(\pi x) + 0.01. \tag{4.45}$$

Several SBP CPR methods with $N = 20$ equally spaced elements of order $p = 7$ and correction terms for the divergence (and restriction, if mentioned) are used as semidiscretisation. The classical Runge-Kutta method of fourth order with $10,000$ equal time steps is used to obtain the discrete solution in the time interval $[0, 3]$.

Results for the SBP CPR method with Lobatto-Legendre basis points and associated quadrature as discrete norm are shown in Figure 4.1. Since the correction for restriction to the boundary is zero, only a correction term for the divergence is used. On the left-hand side, the solution $u(3) = u(3, \cdot)$ obtained with an energy conservative (4.37), local Lax-Friedrichs (4.39) and Osher's (4.40) numerical flux is plotted. On the right-hand side, the evolution of associated discrete momentum $\underline{1}^T \underline{\underline{M}}\, \underline{u}$ and energy $\underline{u}^T \underline{\underline{M}}\, \underline{u}$ in the time interval $[0, 3]$ is visualized.

The ECON flux yields conservation of discrete momentum and energy relative to the initial values of order 10^{-5}, as expected. Using a more accurate time integrator would result in better preservation of these values. Due to the discontinuity around $x = 1$, the results obtained by the ECON flux are not physically relevant and highly oscillatory.

Both the local Lax-Friedrichs and Osher's flux yield good results. After the development of the shock before $t = 0.5$, discrete momentum and energy are constant. Afterwards, momentum is conserved but energy is dissipated, as it is an entropy for Burgers' equation. Around the shock, oscillations develop but remain bounded and the total scheme is stable.

The results in Figure 4.2 are qualitatively similar to those mentioned above. There, a Gauß-Legendre basis is used in an SBP CPR method with correction terms for both divergence and restriction to the boundary. The plots look very similar to those of figure 4.1, but higher accuracy of Gauß-Legendre integration yields slightly less oscillatory solutions for the local Lax-Friedrichs and Osher's flux and a smoother decay of entropy. As before, ECON flux does not yield a physically relevant solution.

In contrast, Figure 4.3 shows results for a Gauß-Legendre basis without the correction term for restriction. In accordance with the theoretical investigations, conservation and stability cannot be guaranteed. A blow-up of energy for the ECON flux occurs around $t = 0.43$. The other solutions are not physically relevant as well, since momentum is lost. Therefore, the additional correction term is necessary.

The results for Roe's flux are shown in Figure 4.4. Stability cannot be guaranteed by using this flux and accordingly the solution obtained by a Lobatto-Legendre basis blows up around $t = 2.5$. The computations using Gauß-Legendre basis remain stable and energy is dissipated, but they do not seem to be as acceptable as those obtained using Osher's or the local Lax-Friedrichs flux. Without the additional correction term for restriction to the boundary, momentum conservation is lost and severe oscillations occur.

Finally, results for high order methods using polynomials of degree $p = 25$ and $p = 50$ are shown in figures 4.5 and 4.6, respectively. The remaining parameters are the same as mentioned above, the

only difference occurs in the increased number (50,000 and 100,000, respectively) of time steps. Of course, very strong oscillations occur, but the method remains stable and conservative. The plots for momentum and energy look precisely like the ones obtained for $p = 7$ and are consequently omitted. These numerical results confirm the proven stability and conservation results even in the case of very high order methods and discontinuous solutions.

4.5. Extension of the CPR idea

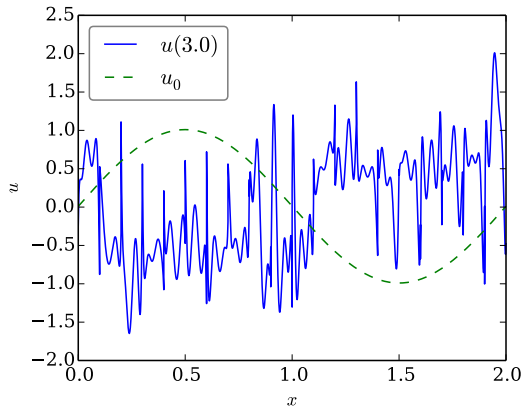
Extending the idea to use a different norm for proving stability does not seem to extend to the corrected formulation of Burgers' equation, at least in a straightforward way. Indeed, multiplying by $\underline{u}^T(\underline{\underline{M}} + \underline{\underline{K}})$ instead of $\underline{u}^T\underline{\underline{M}}$, equation (4.25) becomes

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|^2_{M+K} + \frac{1}{3}\underline{u}^T(\underline{\underline{M}} + \underline{\underline{K}})\underline{\underline{D}}\,\underline{u^2} + \frac{1}{3}\underline{u}^T(\underline{\underline{M}} + \underline{\underline{K}})\underline{u}\,\underline{\underline{D}}\,\underline{u} + \underline{u}^T(\underline{\underline{M}} + \underline{\underline{K}})\underline{\underline{C}}\,(\dots) = 0. \tag{4.46}$$

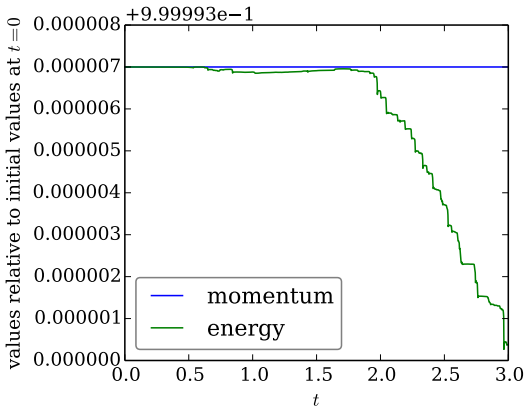
The standard choice $\underline{\underline{C}} = (\underline{\underline{M}} + \underline{\underline{K}})^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$ leads to additional terms

$$\frac{1}{3}\underline{u}^T\underline{\underline{K}}\,\underline{\underline{D}}\,\underline{u}\,\underline{u} + \frac{1}{3}\underline{u}^T\underline{\underline{K}}\,\underline{u}\,\underline{\underline{D}}\,\underline{u} = \frac{1}{3}\underline{u}^T(\underline{\underline{K}}\,\underline{\underline{D}}\,\underline{u} + \underline{\underline{K}}\,\underline{u}\,\underline{\underline{D}})\underline{u} \tag{4.47}$$

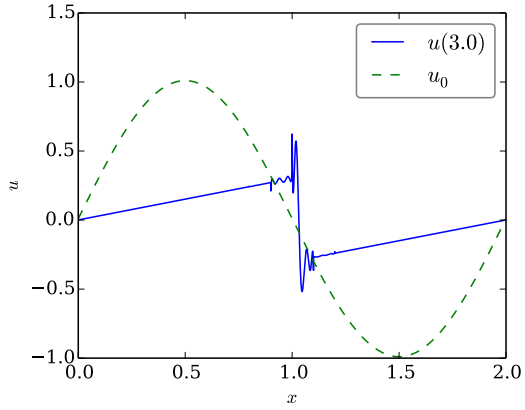
in comparison with the results for $\underline{\underline{K}} = 0$. Enforcing stability by requiring $\underline{\underline{K}}\,\underline{\underline{D}}\,\underline{u} + \underline{\underline{K}}\,\underline{u}\,\underline{\underline{D}}$ to be skew-symmetric (leading to no further contribution) or symmetric (leading to negative contributions for the rate of change $\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|^2_{M+K}$ in the positive definite case) implies $\underline{\underline{K}} = 0$, at least for Gauß-Legendre and Lobatto-Legendre bases of small degree. For brevity, these calculations are not repeated here.



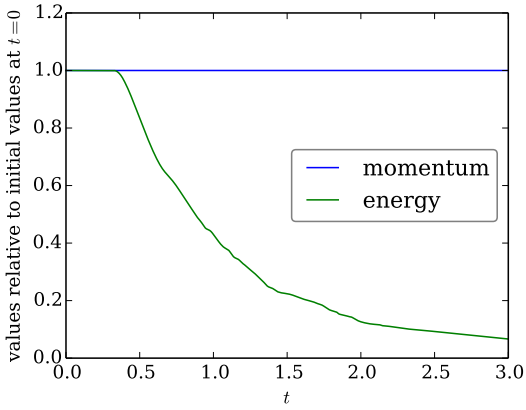
(a) Energy conservative flux.



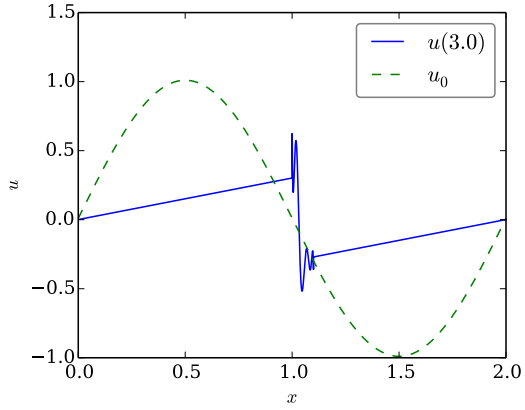
(b) Energy conservative flux.



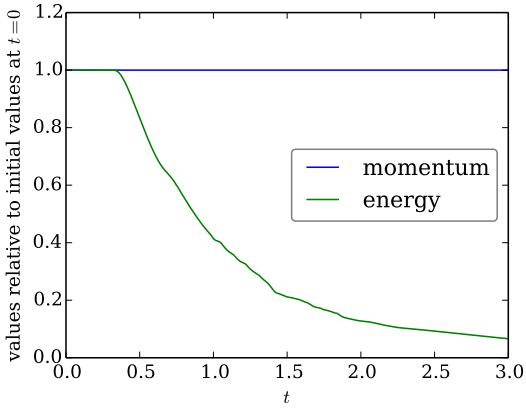
(c) Local Lax-Friedrichs flux.



(d) Local Lax-Friedrichs flux.

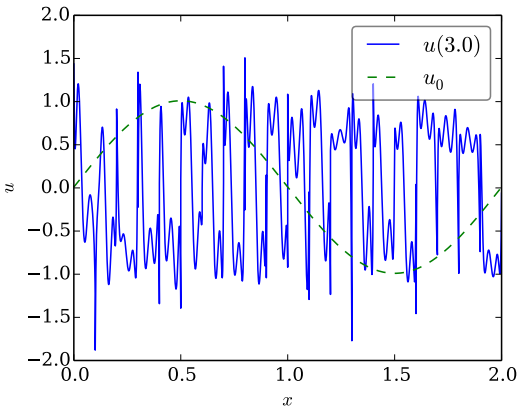


(e) Osher's flux.

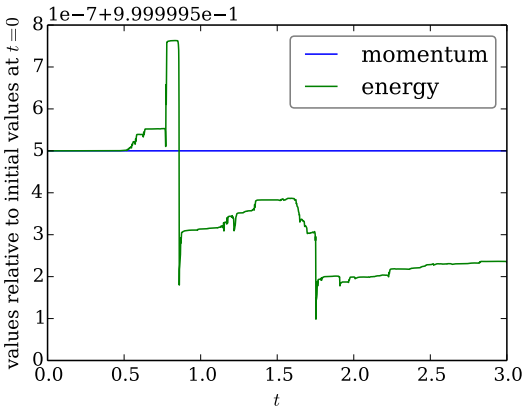


(f) Osher's flux.

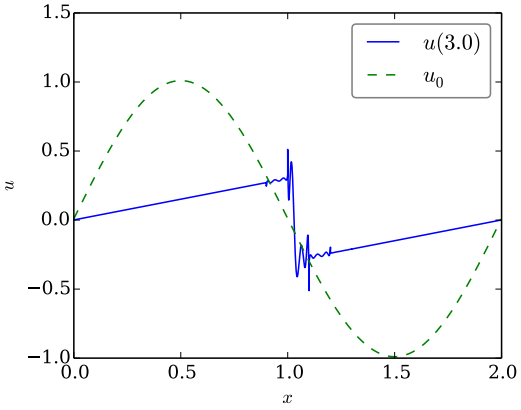
Figure 4.1.: Results of the simulations for Burgers' equation using an SBP CPR method with correction term for the divergence, 20 elements with a *Lobatto-Legendre* basis of order 7 and variant numerical fluxes. On the left-hand side, the values of $u(3)$ (blue) and $u(0) = u_0$ (green) are shown. On the right-hand site, the discrete momentum $\underline{1}^T \underline{\underline{M}}\, \underline{u}$ (blue) and discrete energy $\underline{u}^T \underline{\underline{M}}\, \underline{u}$ (green) in the time interval $[0, 3]$ are plotted.



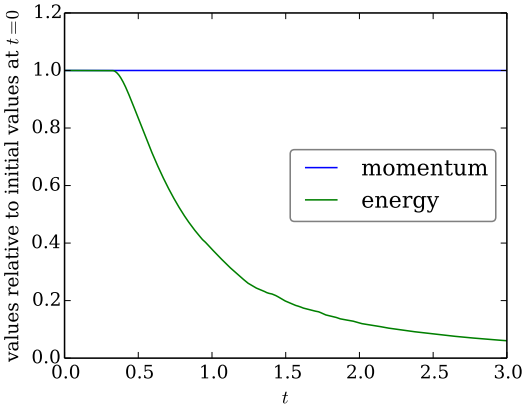
(a) Energy conservative flux.



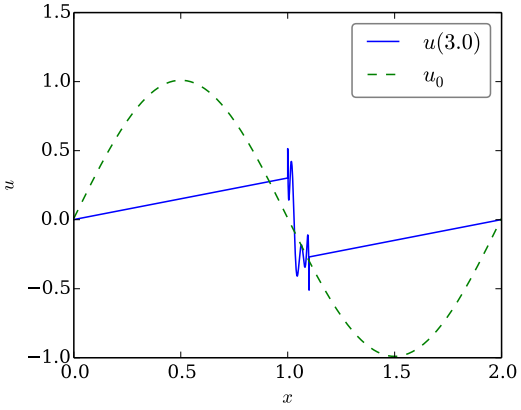
(b) Energy conservative flux.



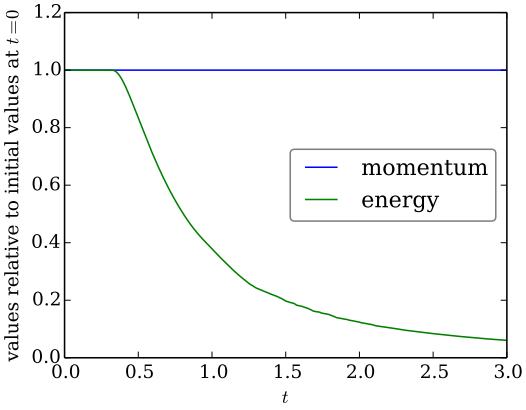
(c) Local Lax-Friedrichs flux.



(d) Local Lax-Friedrichs flux.

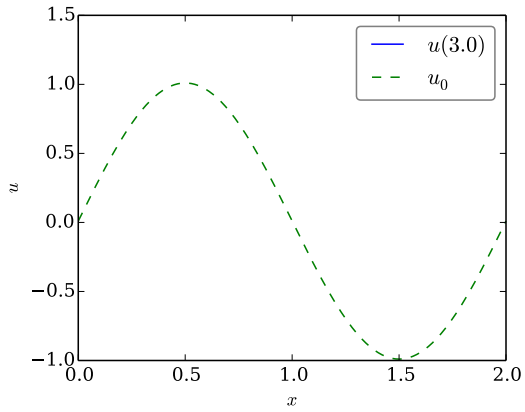


(e) Osher's flux.

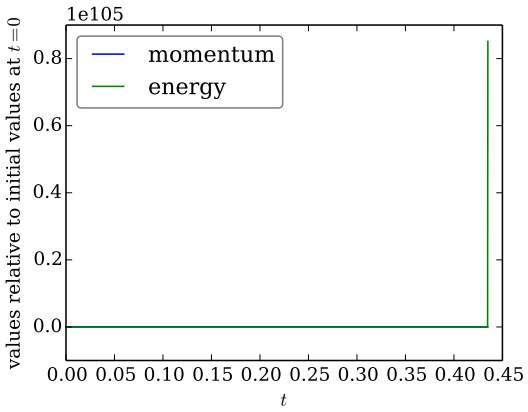


(f) Osher's flux.

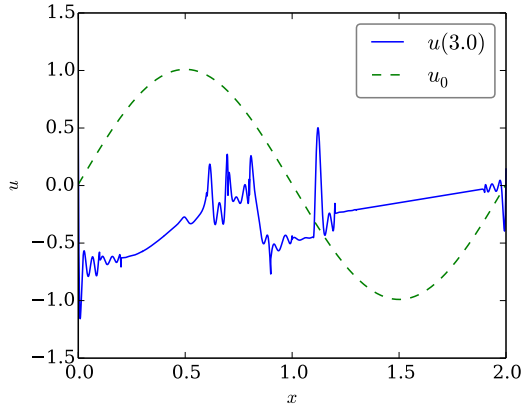
Figure 4.2.: Results of the simulations for Burgers' equation using an SBP CPR method with correction terms for both the divergence and restriction to the boundary, 20 elements with a *Gauß-Legendre* basis of order 7 and variant numerical fluxes. On the left-hand side, the values of $u(3)$ (blue) and $u(0) = u_0$ (green) are shown. On the right-hand site, the discrete momentum $\underline{1}^T \underline{\underline{M}}\, \underline{u}$ (blue) and discrete energy $\underline{u}^T \underline{\underline{M}}\, \underline{u}$ (green) in the time interval $[0, 3]$ are plotted.



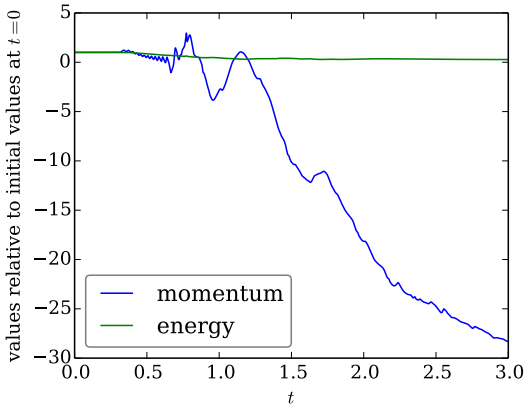
(a) Energy conservative flux.



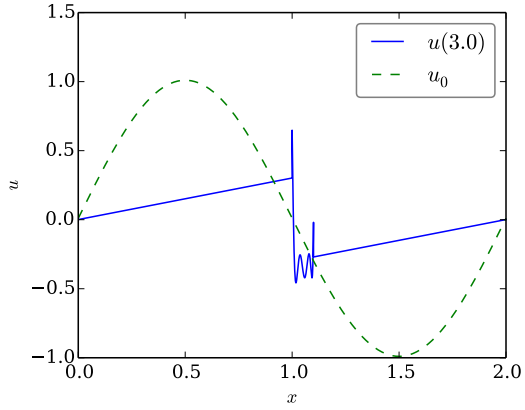
(b) Energy conservative flux.



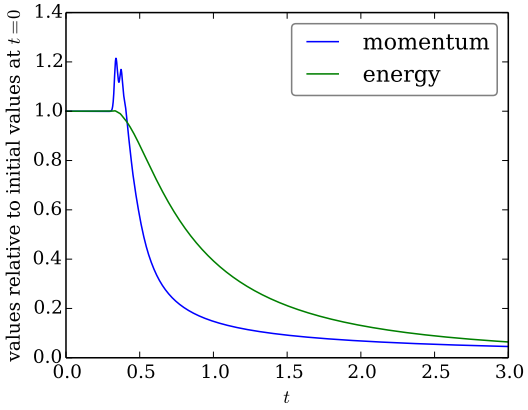
(c) Local Lax-Friedrichs flux.



(d) Local Lax-Friedrichs flux.

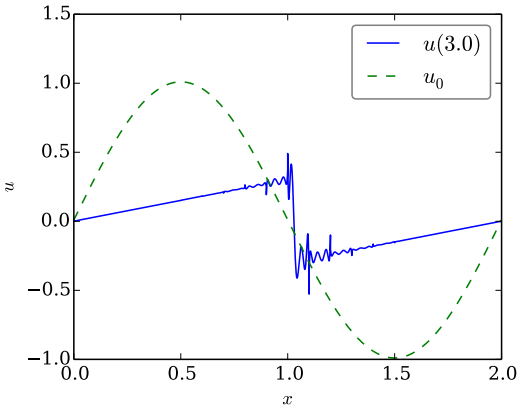


(e) Osher's flux.

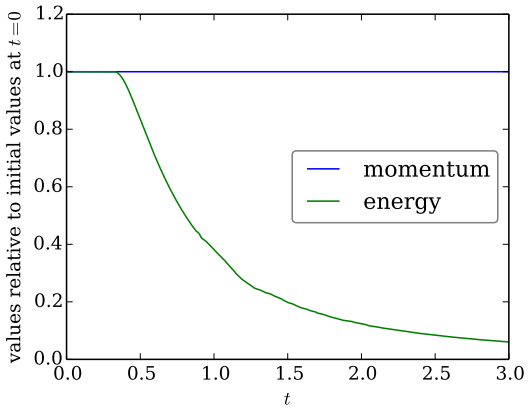


(f) Osher's flux.

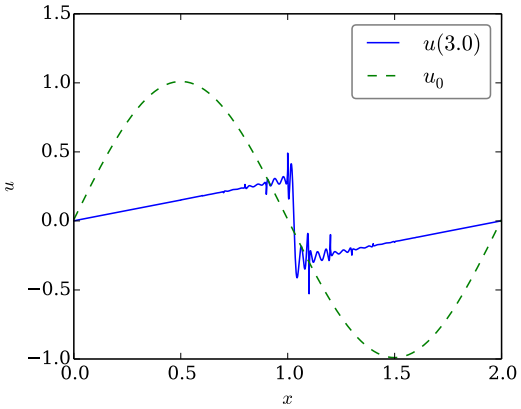
Figure 4.3.: Results of the simulations for Burgers' equation using an SBP CPR method with correction term only for the divergence, 20 elements with a *Gauß-Legendre* basis of order 7 and variant numerical fluxes. On the left-hand side, the values of $u(3)$ (blue) and $u(0) = u_0$ (green) are shown. On the right-hand site, the discrete momentum $\underline{1}^T \underline{\underline{M}}\, \underline{u}$ (blue) and discrete energy $\underline{u}^T \underline{\underline{M}}\, \underline{u}$ (green) in the time interval $[0, 3]$ are plotted.



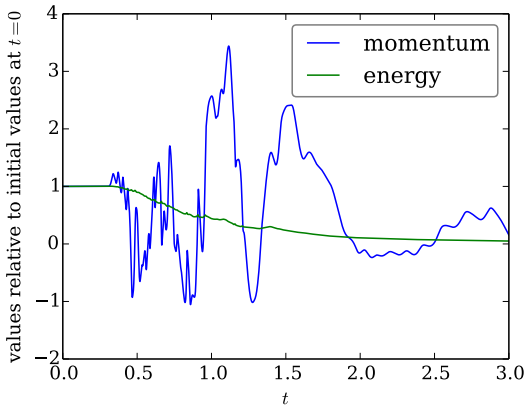
(a) Gauß-Legendre with both correction terms.



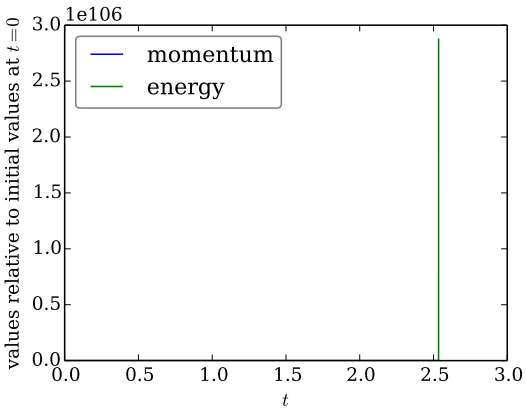
(b) Gauß-Legendre with both correction terms.



(c) Gauß-Legendre with divergence correction.

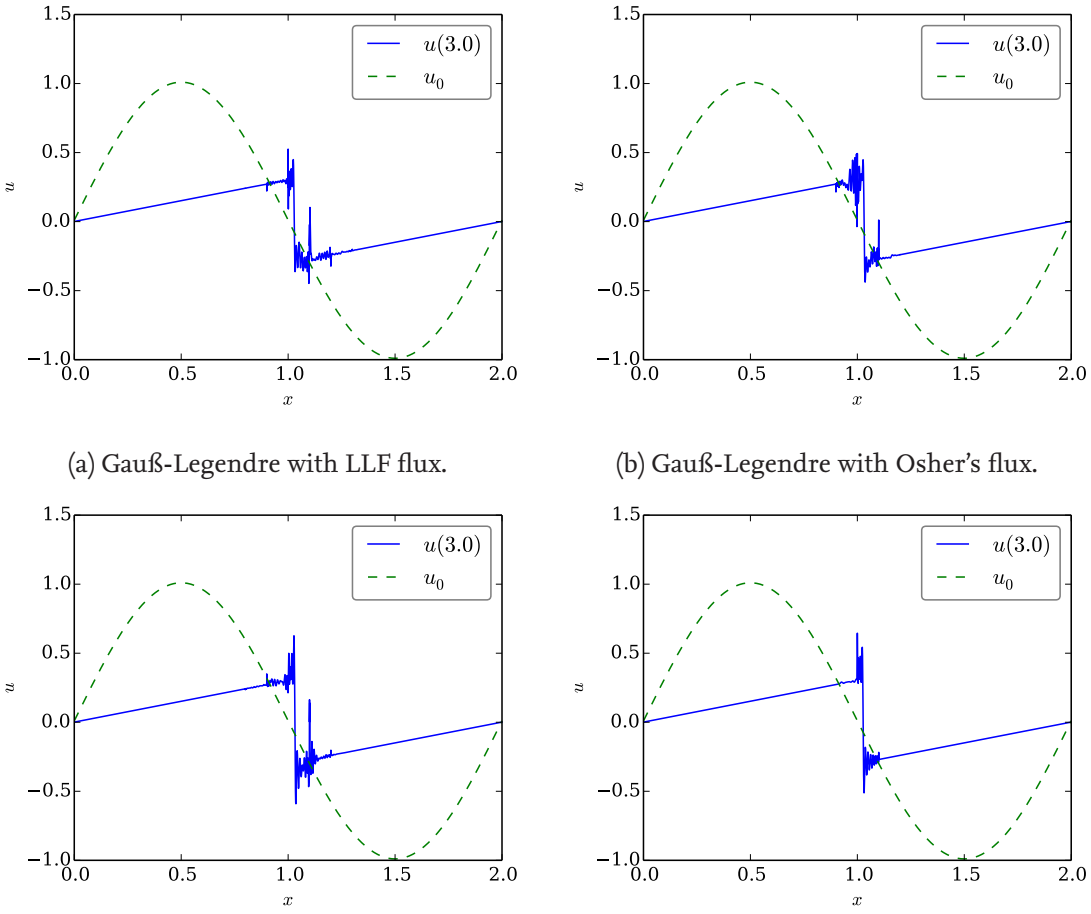


(d) Gauß-Legendre with divergence correction.



(e) Lobatto-Legendre with divergence correction.

Figure 4.4.: Results for Burgers' equation using SBP CPR methods with 20 elements, different bases of order 7 and Roe's flux. In the first row, results for Gauß-Legendre nodes with both correction terms are shown. For the second row, only a divergence correction is used. Finally, (e) presents results for the Lobatto-Legendre basis with correction for the divergence (the restriction correction is zero). In (a) and (c), the values of $u(3)$ (blue) and $u(0) = u_0$ (green) are shown. The other plots visualise the discrete momentum $\underline{1}^T \underline{\underline{M}}\, \underline{u}$ (blue) and discrete energy $\underline{u}^T \underline{\underline{M}}\, \underline{u}$ (green) in the time interval $[0, 3]$.



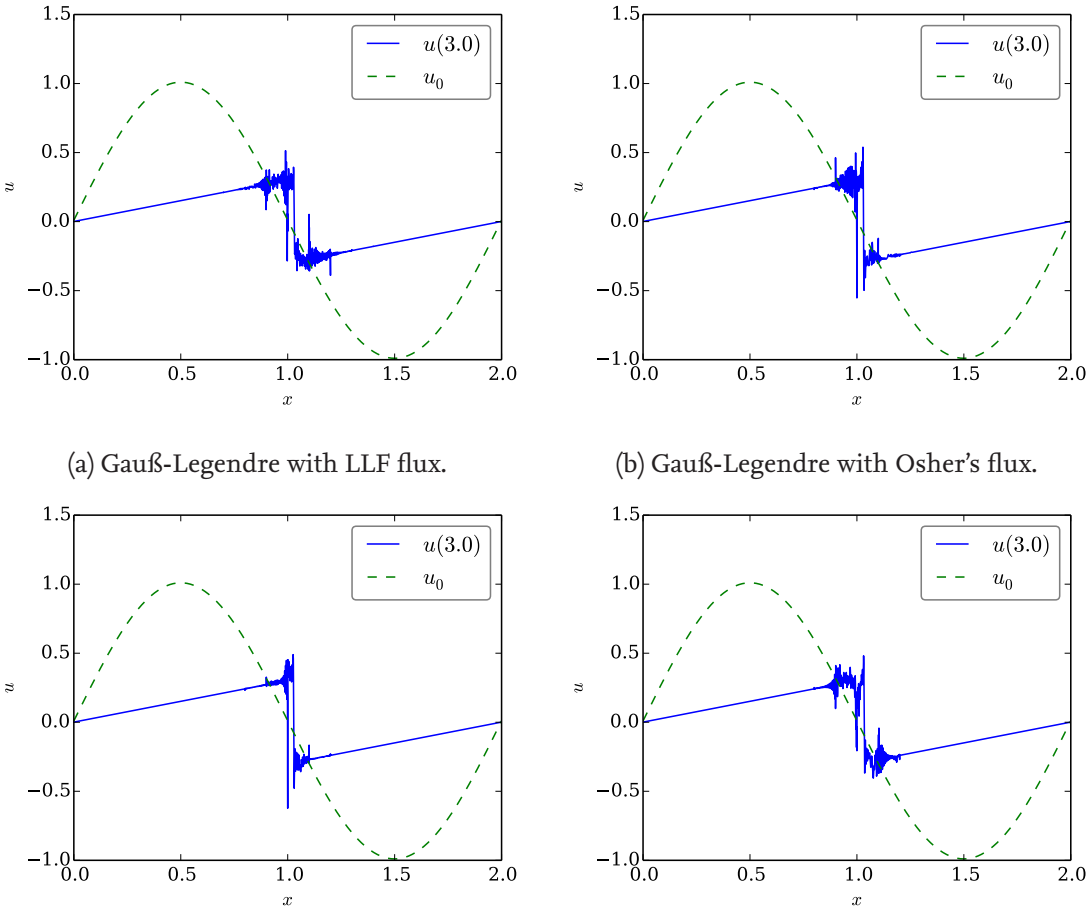
(a) Gauß-Legendre with LLF flux.

(b) Gauß-Legendre with Osher's flux.

(c) Lobatto-Legendre with LLF flux.

(d) Lobatto-Legendre with Osher's flux.

Figure 4.5.: Results of the simulations for Burgers' equation using SBP CPR methods with 20 elements, different bases of order 25 and local Lax-Friedrichs (LLF) or Osher's flux (on the left- and right-hand side, respectively). In the first row, results for the Gauß-Legendre nodes with correction terms for both divergence and restriction are shown. For the second row, a Lobatto-Legendre basis with a correction for the divergence is used. Each Figure shows the values of $u(3)$ (blue) and $u(0) = u_0$ (green).



(a) Gauß-Legendre with LLF flux.

(b) Gauß-Legendre with Osher's flux.

(c) Lobatto-Legendre with LLF flux.

(d) Lobatto-Legendre with Osher's flux.

Figure 4.6.: Results of the simulations for Burgers' equation using SBP CPR methods with 20 elements, different bases of order 50 and local Lax-Friedrichs (LLF) or Osher's flux (on the left- and right-hand side, respectively). In the first row, results for the Gauß-Legendre nodes with correction terms for both divergence and restriction are shown. For the second row, a Lobatto-Legendre basis with a correction for the divergence is used. Each Figure shows the values of $u(3)$ (blue) and $u(0) = u_0$ (green).

5 Abstract view and generalisation

The basic setting described in section 3.1 uses diagonal norm SBP operators and associated quadrature rules with positive weights. These operators have been used in the previous chapters in the context of CPR methods to obtain conservative and stable semidiscretisations for linear advection (3.13) and Burgers' equation (4.1). This chapter provides a more abstract view on the results and more general schemes.

This chapter has been published by order of Professor Sonar [Ranocha et al., 2015a].

5.1. Analytical setting in one dimension

Continuing the investigations of the previous chapters, an analytical setting in the one-dimensional standard element Ω is presented at first. The semidiscretisation in space consists of the representation of a numerical solution in a (real) finite dimensional Hilbert space X_V. Hitherto, X_V has been the space of polynomials of degree $\leq p$, i.e. $\dim X_V = p + 1$. X_V is equipped with a suitable basis $\mathcal{B}_V$, e.g. a Lagrange (interpolation) basis for Gauß-Legendre or Lobatto-Legendre quadrature nodes. With regard to $\mathcal{B}_V$, the *scalar product and associated norm* on X_V are given by a symmetric and positive-definite matrix $\underline{\underline{M}}$, approximating the L^2 norm on X_V, i.e.

$$\underline{u}^T \underline{\underline{M}}\, \underline{v} = \langle \underline{u}, \underline{v} \rangle_M \approx \int_\Omega uv = \langle u, v \rangle_{L^2} . \tag{5.1}$$

In one dimension, a *divergence* (derivative) operator mapping X_V to X_V is represented by a matrix $\underline{\underline{D}}$.

Besides X_V, the vector space of functions on the (one-dimensional) *volume* Ω, a vector space X_B of functions on the (0-dimensional) *boundary* $\partial\Omega$ of the standard element Ω with its associated basis $\mathcal{B}_B$ has to be considered. In the simple one-dimensional case, X_B is a two-dimensional vector space and $\mathcal{B}_B$ is chosen to represent point values at -1 and 1. On the boundary, a bilinear form is represented by a matrix $\underline{\underline{B}}$, approximating the *boundary (surface) integral* in the outward normal direction, i.e. evaluation at the boundary. More precisely, $\underline{\underline{B}}$ maps $X_B \times X_B$ to $\mathbb{R}$ and

$$\underline{u}_B^T \underline{\underline{B}}\, \underline{f}_B = B(u_B, f_B) \approx u_B\, f_B \Big|_{-1}^{1} . \tag{5.2}$$

In the simple one-dimensional setting, u_B and f_B are both scalar functions and $\int_{\partial\Omega} u_B\, f_B \cdot n = u(1)f(1) - u(-1)f(-1)$, i.e. $\underline{\underline{B}} = \operatorname{diag}(-1, 1)$ if $\mathcal{B}_B$ is ordered such that the value at -1 is the first coefficient. With regard to the chosen bases $\mathcal{B}_V$ and $\mathcal{B}_B$, a *restriction* operator is represented by a matrix $\underline{\underline{R}}$, mapping a function u on the volume to its values at the boundary. The SBP property mimics integration by parts and requires

$$\underline{\underline{M}}\, \underline{\underline{D}} + \underline{\underline{D}}^T \underline{\underline{M}} = \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}} . \tag{5.3}$$

A CPR method is further parametrised by a *correction* or *penalty* operator, represented by a matrix $\underline{\underline{C}}$ adapted to the chosen bases. The canonical choice is $\underline{\underline{C}} = \underline{\underline{M}}^{-1} \underline{\underline{R}}^T \underline{\underline{B}}$ as in the previous chapters,

especially for Burgers' equation. For linear advection, other choices of $\underline{\underline{C}}$ are possible, recovering the full range of linearly stable schemes presented by Vincent et al. [2015].

Since nonlinear fluxes $f(u)$ appear and are of interest, nonlinear operations on X_V have to be described. In general, if X_V is a finite-dimensional vector space of polynomials containing polynomials of degree $\leq p$ ($p \geq 1$ and p is minimal), then the product of $u, v \in X_V$ is a polynomial of degree $\leq 2p$, i.e. not in X_V in general. Thus, as described in the paragraph after Lemma 4.2, discrete multiplication is not exact. Multiplying $v \in X_V$ with $u \in X_V$ yields $\underline{\underline{u}}^+ \underline{v} \in X_V^+$, where $X_V^+ \supsetneq X_V$ is a vector space of higher dimension. After this exact multiplication, a *projection* on X_V is performed, resulting in $\underline{\underline{u}}\,\underline{v} \in X_V$.

For a nodal basis $\mathcal{B}_V$, the natural projection is given by pointwise evaluation at the nodes as used in the previous chapters. However, for a modal basis of Legendre polynomials, the natural projection is an L^2 orthogonal projection on X_V. However, this concept does not easily extend to division, since L^2 projection of rational functions is not a simple task.

5.2. Revisiting Burgers' equation

Investigating again a skew-symmetric SBP CPR method without the assumption of a nodal and/or orthogonal basis, some further complications arise. In contrast to the manipulations used to prove Theorem 4.4, $\underline{\underline{u}}$ and $\underline{\underline{M}}$ might not commute, either because the nodal basis is not orthogonal or because a modal basis is chosen. Therefore, the correction terms for the divergence and restriction

$$\underline{c}_{div} = \frac{1}{3}\left(\underline{\underline{u}}\,\underline{\underline{D}}\,\underline{u} - \frac{1}{2}\underline{\underline{D}}\,\underline{\underline{u}}\,\underline{u}\right), \qquad \underline{c}_{res} = \frac{1}{6}\left((\underline{\underline{R}}\,\underline{u})^2 - \underline{\underline{R}}\,\underline{\underline{u}}\,\underline{u}\right), \tag{(4.29)}$$

do not suffice to prove conservation and stability. The reason is again inexactness of discrete multiplication. A multiplication operator $\underline{\underline{u}}$ should be self-adjoint, at least in a finite-dimensional space (and in general, if a correct domain is chosen). Thus, instead of $\underline{\underline{u}}$ in the first term of $\underline{c}_{div}$, the adjoint $\underline{\underline{u}}^*$ of $\underline{\underline{u}}$ with respect to the scalar product induced by $\underline{\underline{M}}$ is proposed. The symmetry condition

$$\left\langle \underline{v}, \underline{\underline{u}}\,\underline{w} \right\rangle_M = \left\langle \underline{\underline{u}}^*\underline{v}, \underline{w} \right\rangle_M \tag{5.4}$$

can be written as

$$\underline{v}^T \underline{\underline{M}}\,\underline{\underline{u}}\,\underline{w} = \underline{v}^T (\underline{\underline{u}}^*)^T \underline{\underline{M}}\,\underline{w}. \tag{5.5}$$

Thus, since $\underline{v}$ and $\underline{w}$ are arbitrary, $\underline{\underline{M}}\,\underline{\underline{u}} = (\underline{\underline{u}}^*)^T \underline{\underline{M}}$, i.e. $\underline{\underline{u}}^* = \underline{\underline{M}}^{-1}\underline{\underline{u}}^T\underline{\underline{M}}$, and the generalised correction terms are

$$\underline{c}_{div} = \frac{1}{3}\left(\underline{\underline{M}}^{-1}\underline{\underline{u}}^T\underline{\underline{M}}\,\underline{\underline{D}}\,\underline{u} - \frac{1}{2}\underline{\underline{D}}\,\underline{\underline{u}}\,\underline{u}\right), \qquad \underline{c}_{res} = \frac{1}{6}\left((\underline{\underline{R}}\,\underline{u})^2 - \underline{\underline{R}}\,\underline{\underline{u}}\,\underline{u}\right). \tag{5.6}$$

Using these correction terms, Theorem 4.4 is generalised by

Theorem 5.1. *If the numerical flux f^{num} satisfies*

$$\frac{1}{6}(u_-^3 - u_+^3) - (u_- - u_+) f^{\text{num}}(u_-, u_+) \leq 0, \tag{5.7}$$

then a general SBP CPR method with $\underline{\underline{C}} = \underline{\underline{M}}^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$ and correction terms (5.6) for both divergence and restriction to the boundary

$$\partial_t \underline{u} + \underline{\underline{D}}\,\frac{1}{2}\underline{u^2} + \underline{c}_{div} + \underline{\underline{C}}\left(\underline{f}^{\text{num}} - \underline{\underline{R}}\,\frac{1}{2}\underline{u^2} - \underline{c}_{res}\right) = 0, \tag{5.8}$$

for the inviscid Burgers' equation (4.1) *is both conservative and stable in the discrete norm* $\|\cdot\|_M$ *induced by* $\underline{\underline{M}}$. *Numerical fluxes fulfilling this condition are inter alia*

- *the energy conservative (ECON) flux* (4.37),
- *the local Lax-Friedrichs (LLF) flux* (4.39),
- *and Osher's flux* (4.40).

Proof. Multiplying $\partial_t \underline{u}$ with $\underline{v}^T \underline{\underline{M}}$, inserting $\underline{\underline{C}} = \underline{\underline{M}}^{-1} \underline{\underline{R}}^T \underline{\underline{B}}$ and applying the SBP property yields

$$\begin{aligned} \underline{v}^T \underline{\underline{M}}\, \partial_t \underline{u} &= -\frac{1}{2} \underline{v}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{\underline{u}}\, \underline{u} - \underline{v}^T \underline{\underline{M}}\, \underline{c}_{div} - \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}} \left(\underline{f}^{\text{num}} - \frac{1}{2} \underline{\underline{R}}\, \underline{\underline{u}}\, \underline{u} - \underline{c}_{res} \right) \\ &= +\frac{1}{2} \underline{v}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{u}}\, \underline{u} - \frac{1}{2} \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}\, \underline{\underline{u}}\, \underline{u} - \underline{v}^T \underline{\underline{M}}\, \underline{c}_{div} - \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}} \left(\underline{f}^{\text{num}} - \frac{1}{2} \underline{\underline{R}}\, \underline{\underline{u}}\, \underline{u} - \underline{c}_{res} \right). \end{aligned} \tag{5.9}$$

Gathering terms and inserting $\underline{c}_{div}$, $\underline{c}_{res}$ from equation (5.6) results in

$$\begin{aligned} \underline{v}^T \underline{\underline{M}}\, \partial_t \underline{u} &= \frac{1}{2} \underline{v}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{u}}\, \underline{u} - \underline{v}^T \underline{\underline{M}}\, \underline{c}_{div} - \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{f}^{\text{num}} + \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{c}_{res} \\ &= \frac{1}{2} \underline{v}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{u}}\, \underline{u} - \frac{1}{3} \underline{v}^T \underline{\underline{u}}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} + \frac{1}{6} \underline{v}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{\underline{u}}\, \underline{u} \\ &\quad - \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{f}^{\text{num}} + \frac{1}{6} \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{\underline{R}}\, \underline{u})^2 - \frac{1}{6} \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}\, \underline{\underline{u}}\, \underline{u}. \end{aligned} \tag{5.10}$$

Applying the SBP property for the third term yields

$$\begin{aligned} \underline{v}^T \underline{\underline{M}}\, \partial_t \underline{u} &= \frac{1}{2} \underline{v}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{u}}\, \underline{u} - \frac{1}{3} \underline{v}^T \underline{\underline{u}}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} + \frac{1}{6} \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}\, \underline{\underline{u}}\, \underline{u} - \frac{1}{6} \underline{v}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{u}}\, \underline{u} \\ &\quad - \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{f}^{\text{num}} + \frac{1}{6} \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{\underline{R}}\, \underline{u})^2 - \frac{1}{6} \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}\, \underline{\underline{u}}\, \underline{u} \\ &= \frac{1}{3} \underline{v}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{\underline{u}}\, \underline{u} - \frac{1}{3} \underline{v}^T \underline{\underline{u}}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} - \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{f}^{\text{num}} + \frac{1}{6} \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{\underline{R}}\, \underline{u})^2. \end{aligned} \tag{5.11}$$

In order to obtain *stability*, $\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \|u\|_M^2 = \underline{u}^T \underline{\underline{M}}\, \partial_t \underline{u}$ has to be considered. Thus, setting $\underline{v} = \underline{u}$ results in

$$\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \|u\|_M^2 = -\underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{f}^{\text{num}} + \frac{1}{6} \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{\underline{R}}\, \underline{u})^2, \tag{5.12}$$

i.e. the same equation as (4.23). Therefore, the proof of Lemma 4.2 can be used to obtain stability.

Investigating *conservation* by setting $\underline{v} = \underline{1}$, using $\underline{\underline{D}}\, \underline{1} = 0$ and $\underline{\underline{u}}\, \underline{1} = \underline{u}$ yields

$$\frac{\mathrm{d}}{\mathrm{d}t} \underline{1}^T \underline{\underline{M}}\, \underline{u} = -\frac{1}{3} \underline{u}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} - \underline{1}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{f}^{\text{num}} + \frac{1}{6} \underline{1}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{\underline{R}}\, \underline{u})^2. \tag{5.13}$$

Rewriting the first term (by the SBP property) as

$$-\frac{1}{3} \underline{u}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} = -\frac{1}{6} \underline{u}^T \underline{\underline{M}}\, \underline{\underline{D}}\, \underline{u} + \frac{1}{6} \underline{u}^T \underline{\underline{D}}^T \underline{\underline{M}}\, \underline{u} - \frac{1}{6} \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}\, \underline{u} = -\frac{1}{6} \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}\, \underline{u} \tag{5.14}$$

results in

$$\frac{\mathrm{d}}{\mathrm{d}t} \underline{1}^T \underline{\underline{M}}\, \underline{u} = -\frac{1}{6} \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}\, \underline{u} - \underline{1}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{f}^{\text{num}} + \frac{1}{6} \underline{1}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{\underline{R}}\, \underline{u})^2. \tag{5.15}$$

As in the proof of Lemma 4.3,

$$\underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\, \underline{\underline{R}}\, \underline{u} = u_R \cdot u_R - u_L \cdot u_L = 1 \cdot u_R^2 - 1 \cdot u_L^2 = \underline{1}^T \underline{\underline{R}}^T \underline{\underline{B}}\, (\underline{\underline{R}}\, \underline{u})^2, \tag{5.16}$$

and the proof can be completed as in Lemma 3.1. □

5.3. Numerical results for dense norm and modal bases

Aside from the basis, the same parameters as in section 4.4 are used to obtain numerical solutions of Burgers' equation (4.1) in the time interval $[0, 3]$. The new nodal bases represent polynomials of degree $\leq p = 7$ using their values at the

- roots $\xi_i = \cos \frac{(2i+1)\pi}{2p+2}$, $i = 0, \dots, p$,
- extrema $\xi_i = \cos \frac{i\pi}{p}$, $i = 0, \dots, p$

of Chebyshev polynomial T_{p+1} of first kind or the

- roots $\xi_i = \cos \frac{(i+1)\pi}{p+2}$, $i = 0, \dots, p$

of the Chebyshev polynomial U_{p+1} of second kind. The differentiation and norm matrices $\underline{\underline{D}}$, $\underline{\underline{M}}$ are computed via their representation for Legendre polynomials and a basis transformation using the associated Vandermonde matrix, see section 3.4 and appendix A. Multiplication is conducted pointwise at the corresponding Chebyshev nodes. For these bases, $\underline{\underline{M}}$ is not diagonal and multiplication operators $\underline{u}$ are not $\underline{\underline{M}}$-self-adjoint in general.

Additionally, a modal basis of Legendre polynomials is used, performing exact multiplication followed by an orthogonal projection. For this orthogonal basis, a multiplication operator $\underline{u}$ is in general not diagonal, but $\underline{\underline{M}}$-self-adjoint, as the following calculation for arbitrary polynomials u, v, w of degree $\leq p$ shows:

$$\left\langle \underline{v}, \underline{\underline{u}}\, \underline{w} \right\rangle_M = \underline{v}^T \underline{\underline{M}}\, \underline{\underline{u}}\, \underline{w} = \int v \operatorname{proj}(u\, w) = \int v\, u\, w = \int \operatorname{proj}(u\, v)\, w = \underline{v}^T \underline{\underline{u}}^T \underline{\underline{M}}\, \underline{w} = \left\langle \underline{\underline{u}}\, \underline{v}, \underline{w} \right\rangle_M . \tag{5.17}$$

The third and fourth equality follow from the orthogonality of Legendre polynomials. Thus, multiplication operators $\underline{u}$ are $\underline{\underline{M}}$-self-adjoint.

An interpolation approach to compute the initial values for a Legendre basis using the nodes of all nodal bases presented in Figure 5.1 has been used. There is no visual difference between results for these different sets of nodes. In the following, interpolation via Gauß-Legendre nodes has been used.

The results of the computations using the local Lax-Friedrichs flux are shown in Figures 5.1 and 5.2. For comparison, the results using Gauß-Legendre and Lobatto-Legendre bases as in the previous chapters are included in the first rows. The values of $u(3)$ are in general similar – two approximately affine-linear parts and a discontinuous part with oscillations around $x = 1$. Despite of this, the intensity of oscillations depends on the bases and associated projection used for multiplication.

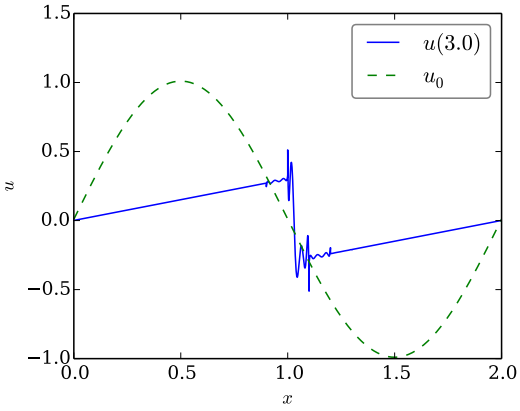
In this case, the roots of Chebyshev polynomials of second kind seem to perform worst, whereas Gauß-Legendre nodes and modal Legendre polynomials seem to be least oscillatory and visually indistinguishable. Contrary, the computations using a nodal basis are much more efficient, since only simple multiplication of nodal values has to be performed.

As expected, momentum is conserved for all bases and the discrete energy (entropy) is constant until $t \approx 0.5$ and decays afterwards, as can be seen in Figure 5.2.

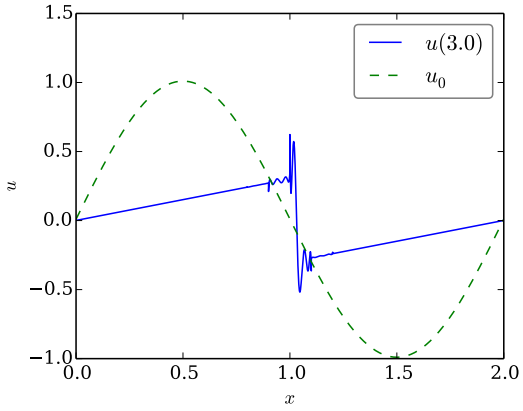
These results are obtained using general SBP CPR methods (5.8) with both correction terms for divergence and restriction (5.6). Ignoring a non-trivial correction term for a nodal basis leads to physically useless results, as can be seen for example in Figure 4.3. Results without the skew-symmetric

correction $\underline{c}_{div}$ are not plotted here. Additionally, the correction term $\underline{c}_{div}$ using the $\underline{\underline{M}}$-adjoint multiplication operator is verified numerically, since using the simple multiplication as in the previous chapter gives erroneous results, again not shown here.

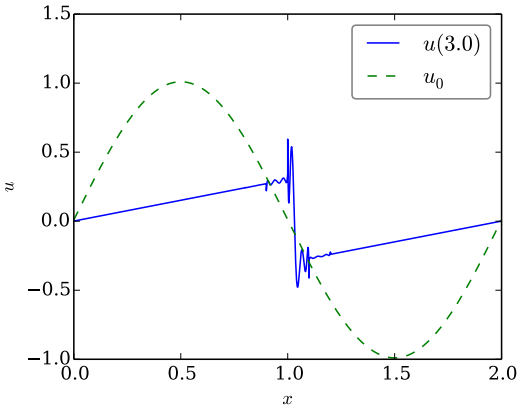
Remarkably, the results (not plotted here) using a modal Legendre basis and either both or no correction term ($\underline{c}_{div}$, $\underline{c}_{res}$) are visually indistinguishable. Additionally, using only $\underline{c}_{res}$ yields the same results. Contrary, using only a correction for the divergence results in varying momentum and physically useless results. Using an exact orthogonal projection during multiplication seems to be a good idea, but an analytical investigation of this phenomenon remains an open problem.



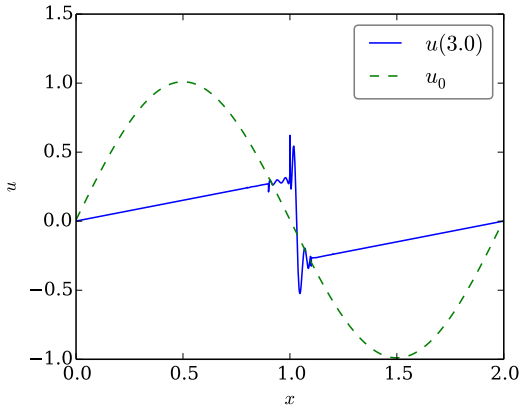
(a) Gauß-Legendre.



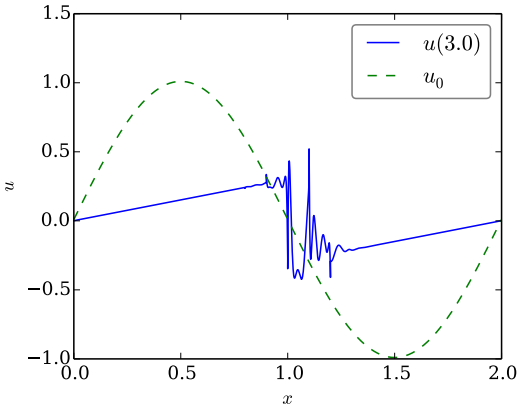
(b) Lobatto-Legendre.



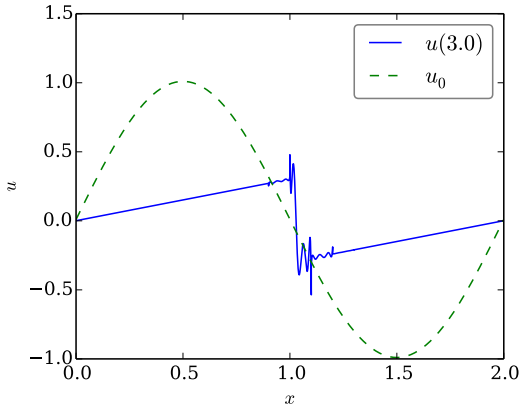
(c) Chebyshev first kind, roots.



(d) Chebyshev first kind, extrema.

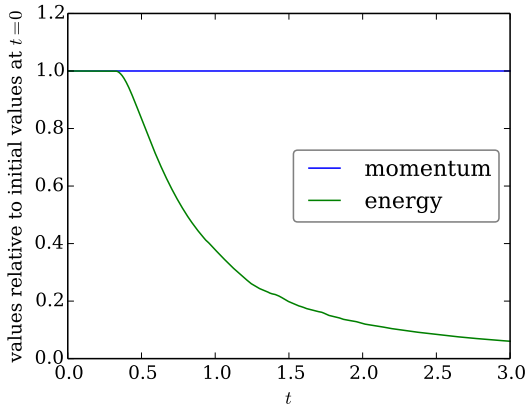


(e) Chebyshev second kind, roots.

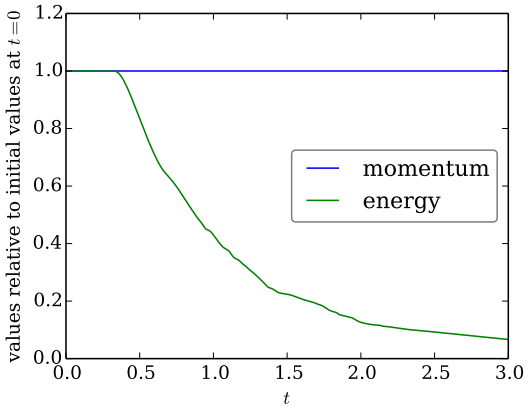


(f) Legendre.

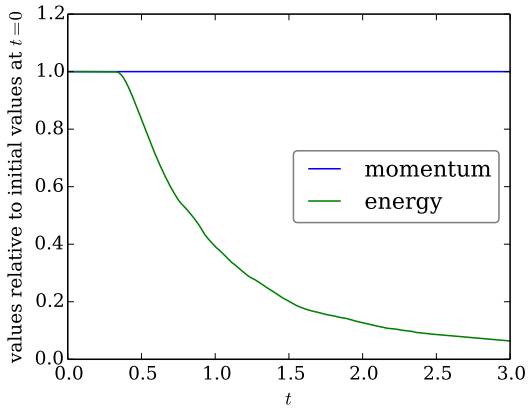
Figure 5.1.: Results of the simulations for Burgers' equation using general SBP CPR methods with 20 elements, different bases of order 7 and local Lax-Friedrichs (LLF) flux. Corrections for both divergence and restriction are used. Each Figure shows the values of $u(3)$ (blue) and $u(0) = u_0$ (green) for different bases.



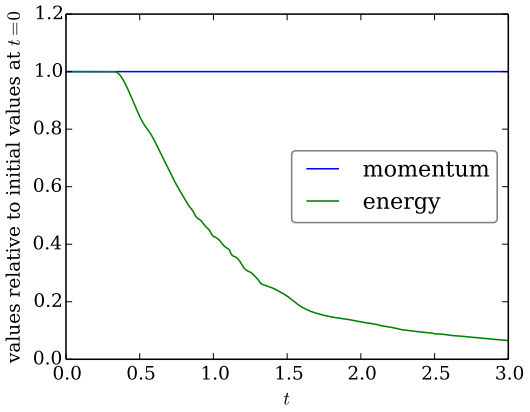
(a) Gauß-Legendre.



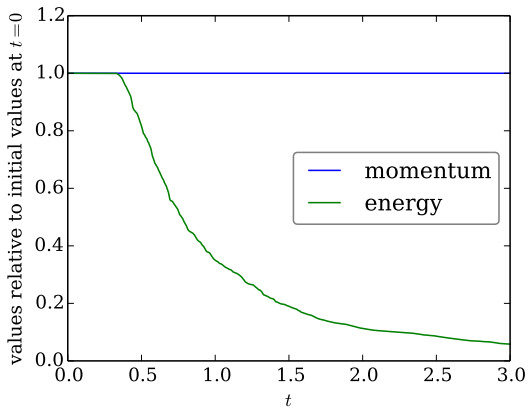
(b) Lobatto-Legendre.



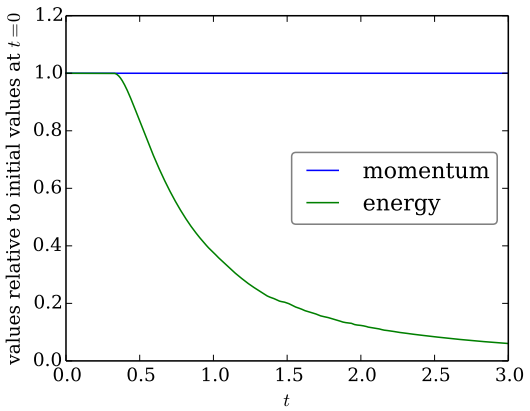
(c) Chebyshev first kind, roots.



(d) Chebyshev first kind, extrema.



(e) Chebyshev second kind, roots.



(f) Legendre.

Figure 5.2.: Results of the simulations for Burgers' equation using general SBP CPR methods with 20 elements, different bases of order 7 and local Lax-Friedrichs (LLF) flux. Corrections for both divergence and restriction are used. Each Figure shows the discrete momentum $\underline{1}^T \underline{\underline{M}}\, \underline{u}$ (blue) and discrete energy $\underline{u}^T \underline{\underline{M}}\, \underline{u}$ (green) for different bases.

5.4. A brief view on a numerical setting

The analytical setting of section 5.1 is based on a given solution space X_V for the one-dimensional standard element, since the investigations in this work started from CPR methods, extending DG methods which are also described by a fundamental basis. Contrary, the theory of SBP operators originates in FD methods, classically not equipped with a solution basis other than the nodal values. Nevertheless, Gassner [2013] adapted the SBP framework to a DGSEM with nodal Lobattto-Legendre basis and lumped mass matrix. Additionally, Fernández et al. [2014a] proposed a generalised SBP framework in one dimension based on nodal values without an analytical basis. Instead, the operators are required to fulfil the SBP property and some accuracy conditions, i.e. they should be exact for polynomials up to some degree $p \geq 1$. These ideas were extended by Hicken et al. [2015] to multi-dimensional operators, focussing on diagonal-norm SBP operators on simplex elements in two and three dimensions, i.e. triangles and tetrahedra.

These extensions were only applied to linear advection with constant velocity and proved to be conservative and stable in the norm associated to the SBP operator. Relaxing accuracy conditions potentially results in additional free parameters, allowing the construction of specialised schemes for different purposes. As already proved by Hicken and Zingg [2013], SBP operators are tightly coupled to quadrature rules. Thus, different quadrature rules can be used to obtain SBP operators and vice versa.

All investigations conducted in the previous chapters and sections directly extend to these generalised FD SBP operators with diagonal or dense norm, respectively. Additionally, since these operators are described by the same matrices used hitherto in the investigations, they can be simply plugged in the numerical method for the calculations – up to the last step. In the analytical setting, the solution is completely determined by the given coefficients with regard to the chosen basis, i.e. sub-cell resolution of arbitrary accuracy is given. Especially, the solution can be plotted exactly as it is used in the computations. Contrary, using only nodal values at a given set of points without an interpretation as coefficients of a known basis, only these point values can be plotted as output seriously. Performing any interpolation would be a guess, but can in general not describe the solution accurately. From the author's point of view, this is a serious drawback of the numerical setting without a basis as foundation. The inability to describe a modal basis does not seem to be equally unfavourable, since computing a correct orthogonal projection for division is not a straightforward task and nodal methods are much more efficient regarding evaluation times for nonlinear operations.

A solution of the interpolation problem would be to construct a basis describing a given SBP operator. For example, Gassner [2013] constructed a basis for a specially chosen FD SBP operator. However, there does not seem to be a straightforward way to construct such a basis in general.

6 Extension to multiple dimensions

Hitherto, all investigations were conducted in one space dimension. Although the formulation used was intended to be general, some changes need to be introduced. Additionally, since boundary integrals in multiple dimensions can not be evaluated as easily as in the one-dimensional case as pointwise evaluations, some further complications arise.

6.1. Analytical setting in multiple dimensions

In this section, an extension of the abstract description given in section 5.1 to multiple dimensions in described. There may be some similarities to the numerical setting proposed by Hicken et al. [2015] for multiple dimensions, but the approach is based on an analytical setting and was developed independently.

All computations are performed after a diffeomorphic mapping to the d-dimensional standard element $\Omega \subset \mathbb{R}^d$. The solution is semidiscretely approximated as a member of a (real) finite-dimensional Hilbert space X_V with basis $\mathcal{B}_V$ in the *volume* Ω, i.e. X_V consists of functions on Ω, e.g. polynomials of degree $\leq p$. As in the one-dimensional case, the scalar product is induced by a symmetric and positive-definite matrix $\underline{\underline{M}}$ (known as *mass*-matrix for DG methods) and approximates the L^2 norm on Ω

$$\underline{u}^T \underline{\underline{M}}\, \underline{v} = \langle \underline{u}, \underline{v} \rangle_M \approx \int_\Omega uv = \langle u, v \rangle_{L^2(\Omega)} . \tag{6.1}$$

Since all computations are performed using coordinates and the aim of this work does not include curvilinear coordinates, it seems to be acceptable not to insist on coordinate free formulations but to adopt standard Cartesian coordinates in $\mathbb{R}^d$. Therefore, the *divergence* operator, mapping X_V^d to X_V is given by d *derivative* operators $\underline{\underline{D}}_i, i \in \{1, \dots, d\}$, representing the partial derivative in the i-th coordinate direction:

$$\underline{\underline{D}} = \left(\underline{\underline{D}}_1, \dots, \underline{\underline{D}}_d \right) . \tag{6.2}$$

Analogously to the one-dimensional case, a Hilbert space X_B with basis $\mathcal{B}_B$ consisting of functions on the *boundary* $\partial\Omega$ of Ω is needed. Contrary to the one-dimensional case, there is no canonical basis, since no finite number of nodal values suffices to describe boundary values of an arbitrary continuous function on Ω. Additionally, the outer normal boundary integration operator is split into two operators. Similarly to X_V, X_B is a Hilbert space with scalar product induced by $\underline{\underline{B}}$, representing an L^2 boundary integral

$$\underline{u}_B^T \underline{\underline{B}}\, \underline{v}_B = \langle \underline{u}_B, \underline{v}_B \rangle_B \approx \int_{\partial\Omega} u_B v_B = \langle u_B, v_B \rangle_{L^2(\partial\Omega)} . \tag{6.3}$$

Additionally, multiplication with the i-th component n_i of the outer unit normal $\underline{n}$ is represented by an operator $\underline{\underline{N}}_i, i \in \{1, \dots, d\}$. In section 5.1, $\underline{\underline{B}}$ denoted a bilinear form performing integration with the outer normal, i.e. $\underline{\underline{B}}$ of section 5.1 is $\underline{\underline{B}}\,\underline{1} = \underline{\underline{B}}\,\underline{\underline{N}}_1$ in this setting. Since exact point values at both end points were used, this reduces to $\underline{\underline{B}}\,\underline{1} = \underline{\underline{B}}\,\underline{\underline{N}}_1 = \underline{\underline{I}}\,\underline{\underline{N}}_1 = \underline{\underline{N}}_1$. Here and in the following, $\underline{\underline{I}}$

denotes the identity matrix for the one-dimensional basis (i.e. of size $(p+1)\times(p+1)$) or of the size indicated by a subscript, if existing.

X_V and X_B are coupled via a *restriction* operator $\underline{\underline{R}}$, representing restriction of a function u on Ω to the boundary $\partial\Omega$, known as trace in the setting of Sobolev spaces. In this setting, the restriction and integral operators would be continuous with domain $H^1(\Omega)$ or $L^2(\partial\Omega)$. Contrary, the derivative operators can be defined as discontinuous mappings on $H^1(\Omega)$. Another possibility would be to define them on $H^1(\Omega)$ but with values in $L^2(\Omega)$.

Finally, the SBP property

$$\underline{\underline{M}}\left(\underline{\underline{D}}_1,\dots,\underline{\underline{D}}_d\right)+\left(\underline{\underline{D}}_1{}^T,\dots,\underline{\underline{D}}_d{}^T\right)(\underline{\underline{I}}_d\otimes\underline{\underline{M}})=\underline{\underline{R}}^T\underline{\underline{B}}\left(\underline{\underline{N}}_1,\dots,\underline{\underline{N}}_d\right)(\underline{\underline{I}}_d\otimes\underline{\underline{R}}) \tag{6.4}$$

is required, mimicking integration by parts in multiple dimensions via the divergence theorem

$$\int_\Omega u\,\mathrm{div}\,f+\int_\Omega \mathrm{grad}\,u\cdot f=\int_{\partial\Omega}u\,f\cdot n \tag{6.5}$$

for a scalar field u, a vector field f and normal domain Ω, regular enough. Here and in the following, $\otimes$ denotes the bilinear *Kronecker product* for matrices $A\in\mathbb{R}^{k\times l}$ and $B\in\mathbb{R}^{m\times n}$

$$A\otimes B:=\begin{pmatrix}A_{11}B & \dots & A_{1l}B\\ \vdots & \ddots & \vdots\\ A_{k1}B & \dots & A_{kl}B\end{pmatrix}\in\mathbb{R}^{km\times ln}. \tag{6.6}$$

Therefore, formulating (6.4) column by column

$$\underline{\underline{M}}\,\underline{\underline{D}}_i+\underline{\underline{D}}_i^T\underline{\underline{M}}=\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}}_i\underline{\underline{R}},\quad i\in\{1,\dots,d\}. \tag{6.7}$$

The straightforward extension of a one-dimensional basis to a multi-dimensional basis can be conducted using a tensor product structure. Thus, if ϕ_i, $i\in\{0,\dots,p\}$, are one-dimensional basis functions, the d-dimensional basis consists of $\phi_{i_1}(x_1)\dots\phi_{i_d}(x_d)$, $i_1,\dots,i_d\in\{0,\dots,p\}$. Using Fortran ordering for vectorisation, the coefficients are sorted with the index for the first coordinate x_1 varying fastest

$$\underline{u}=(u_{0,0,\dots,0},u_{1,0,\dots,0},\dots,u_{p,0,\dots,0},u_{0,1,0,\dots,0},u_{1,1,0,\dots,0},\dots,u_{p,1,0,\dots,0},\dots,u_{0,p,\dots,p},\dots,u_{p,p,\dots,p})^T. \tag{6.8}$$

Denoting the one-dimensional matrices with small letters $\underline{\underline{m}}$, $\underline{\underline{r}}$, $\underline{\underline{b}}\,(=\underline{\underline{I}})$, $\underline{\underline{n}}$ and $\underline{\underline{d}}$, the tensor product matrices are

$$\underline{\underline{M}}=\underline{\underline{m}}\otimes\cdots\otimes\underline{\underline{m}}, \tag{6.9a}$$

$$\underline{\underline{R}}=\begin{pmatrix}\underline{\underline{r}}\otimes\underline{\underline{I}}\otimes\cdots\otimes\underline{\underline{I}}\\ \underline{\underline{P}}_2(\underline{\underline{I}}\otimes\underline{\underline{r}}\otimes\underline{\underline{I}}\otimes\cdots\otimes\underline{\underline{I}})\\ \vdots\\ \underline{\underline{P}}_d(\underline{\underline{I}}\otimes\cdots\otimes\underline{\underline{I}}\otimes\underline{\underline{r}})\end{pmatrix}, \tag{6.9b}$$

$$\underline{\underline{B}}=\underline{\underline{I}}_{2d}\otimes\underline{\underline{m}}\otimes\cdots\otimes\underline{\underline{m}}, \tag{6.9c}$$

$$\underline{\underline{N}}_1 = \operatorname{diag}\Big(0,\dots,0,\underline{\underline{n}} \otimes \underline{\underline{I}}_{(p+1)^d}\Big),\ \dots,\ \underline{\underline{N}}_d = \operatorname{diag}\Big(\underline{\underline{n}} \otimes \underline{\underline{I}}_{(p+1)^d},0,\dots,0\Big), \tag{6.9d}$$

$$\underline{\underline{D}}_1 = \underline{\underline{I}} \otimes \cdots \otimes \underline{\underline{I}} \otimes \underline{\underline{d}},\ \dots,\ \underline{\underline{D}}_d = \underline{\underline{d}} \otimes \underline{\underline{I}} \otimes \cdots \otimes \underline{\underline{I}}. \tag{6.9e}$$

Here, d matrices per Kronecker product are used. $\underline{\underline{P}}_2,\dots,\underline{\underline{P}}_d$ are permutation matrices sorting the values at the faces of the cube $[-1,1]^d$ (using again Fortran ordering)

$$\underline{\underline{R}}\,\underline{u} = (u_{0,\dots,0,0},\dots,u_{p,\dots,p,0},u_{0,\dots,0,p},\dots,u_{p,\dots,p,p},\dots,u_{0,0\dots,0},\dots,u_{0,p,\dots,p},u_{p,0\dots,0},\dots,u_{p,p,\dots,p})^T. \tag{6.10}$$

In order to get a more concise and clear representation, the following computations are restricted to two dimensions $d = 2$. In order to verify the SBP property (6.7) in y-direction, i.e. $i = 2$,

$$\begin{aligned}\underline{\underline{M}}\,\underline{\underline{D}}_2 + \underline{\underline{D}}_2^T\underline{\underline{M}} &= (\underline{\underline{m}} \otimes \underline{\underline{m}})(\underline{\underline{d}} \otimes \underline{\underline{I}}) + (\underline{\underline{d}}^T \otimes \underline{\underline{I}})(\underline{\underline{m}} \otimes \underline{\underline{m}}) \\ &= (\underline{\underline{m}}\,\underline{\underline{d}}) \otimes \underline{\underline{m}} + (\underline{\underline{d}}^T\underline{\underline{m}}) \otimes \underline{\underline{m}} = (\underline{\underline{m}}\,\underline{\underline{d}} + \underline{\underline{d}}^T\underline{\underline{m}}) \otimes \underline{\underline{m}}\end{aligned} \tag{6.11}$$

and (bearing $\underline{\underline{b}} = \underline{\underline{I}}$ in mind)

$$\begin{aligned}\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}}_2\underline{\underline{R}} &= \Big(\underline{\underline{r}}^T \otimes \underline{\underline{I}}, (\underline{\underline{I}} \otimes \underline{\underline{r}}^T)\underline{\underline{P}}_2^T\Big)\begin{pmatrix}\underline{\underline{I}}_2 \otimes \underline{\underline{m}} & 0\\ 0 & \underline{\underline{I}}_2 \otimes \underline{\underline{m}}\end{pmatrix}\begin{pmatrix}\underline{\underline{n}} \otimes \underline{\underline{I}} & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\underline{\underline{r}} \otimes \underline{\underline{I}}\\ \underline{\underline{P}}_2(\underline{\underline{I}} \otimes \underline{\underline{r}})\end{pmatrix} \\ &= (\underline{\underline{r}}^T \otimes \underline{\underline{I}})(\underline{\underline{n}} \otimes \underline{\underline{m}})(\underline{\underline{r}} \otimes \underline{\underline{I}}) = (\underline{\underline{r}}^T\underline{\underline{n}}\,\underline{\underline{r}}) \otimes \underline{\underline{m}} = (\underline{\underline{r}}^T\underline{\underline{b}}\,\underline{\underline{n}}\,\underline{\underline{r}}) \otimes \underline{\underline{m}}\end{aligned} \tag{6.12}$$

need to be equal. Indeed, this is true, since the one-dimensional SBP property

$$\underline{\underline{m}}\,\underline{\underline{d}} + \underline{\underline{d}}^T\underline{\underline{m}} = \underline{\underline{r}}^T\underline{\underline{b}}\,\underline{\underline{n}}\,\underline{\underline{r}} \tag{6.13}$$

is satisfied. In x-direction ($i = 1$), the corresponding terms are

$$\begin{aligned}\underline{\underline{M}}\,\underline{\underline{D}}_1 + \underline{\underline{D}}_1^T\underline{\underline{M}} &= (\underline{\underline{m}} \otimes \underline{\underline{m}})(\underline{\underline{I}} \otimes \underline{\underline{d}}) + (\underline{\underline{I}} \otimes \underline{\underline{d}}^T)(\underline{\underline{m}} \otimes \underline{\underline{m}}) \\ &= \underline{\underline{m}} \otimes (\underline{\underline{m}}\,\underline{\underline{d}}) + \underline{\underline{m}} \otimes (\underline{\underline{d}}^T\underline{\underline{m}}) = \underline{\underline{m}} \otimes (\underline{\underline{m}}\,\underline{\underline{d}} + \underline{\underline{d}}^T\underline{\underline{m}})\end{aligned} \tag{6.14}$$

and (using again $\underline{\underline{b}} = \underline{\underline{I}}$)

$$\begin{aligned}\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}}_1\underline{\underline{R}} &= \Big(\underline{\underline{r}}^T \otimes \underline{\underline{I}}, (\underline{\underline{I}} \otimes \underline{\underline{r}}^T)\underline{\underline{P}}_2^T\Big)\begin{pmatrix}\underline{\underline{I}}_2 \otimes \underline{\underline{m}} & 0\\ 0 & \underline{\underline{I}}_2 \otimes \underline{\underline{m}}\end{pmatrix}\begin{pmatrix}0 & 0\\ 0 & \underline{\underline{n}} \otimes \underline{\underline{I}}\end{pmatrix}\begin{pmatrix}\underline{\underline{r}} \otimes \underline{\underline{I}}\\ \underline{\underline{P}}_2(\underline{\underline{I}} \otimes \underline{\underline{r}})\end{pmatrix} \\ &= (\underline{\underline{I}} \otimes \underline{\underline{r}}^T)\underline{\underline{P}}_2^T(\underline{\underline{n}} \otimes \underline{\underline{m}})\underline{\underline{P}}_2(\underline{\underline{I}} \otimes \underline{\underline{r}}) = (\underline{\underline{I}} \otimes \underline{\underline{r}}^T)(\underline{\underline{m}} \otimes \underline{\underline{n}})(\underline{\underline{I}} \otimes \underline{\underline{r}}) \\ &= \underline{\underline{m}} \otimes (\underline{\underline{r}}^T\underline{\underline{n}}\,\underline{\underline{r}}) = \underline{\underline{m}} \otimes (\underline{\underline{r}}^T\underline{\underline{b}}\,\underline{\underline{n}}\,\underline{\underline{r}}).\end{aligned} \tag{6.15}$$

The third equality is fulfilled by the definition of $\underline{\underline{P}}_2$. Indeed, since

$$(\underline{\underline{I}} \otimes \underline{\underline{r}})\underline{u} = (\underline{\underline{I}} \otimes \underline{\underline{r}})\begin{pmatrix}u_{0,0}\\ \vdots\\ u_{p,0}\\ \vdots\\ u_{0,p}\\ \vdots\\ u_{p,p}\end{pmatrix} = \begin{pmatrix}u_{l,0}\\ u_{r,0}\\ \vdots\\ u_{l,p}\\ u_{r,p}\end{pmatrix}, \tag{6.16}$$

where the indices l and r denote the values of $u_{\cdot,i}$ for fixed index i at the left and right boundaries, and $\underline{\underline{P}}_2$ is defined by

$$\underline{\underline{P}}_2(\underline{\underline{I}} \otimes \underline{\underline{r}})\underline{u} = \begin{pmatrix} u_{l,0} \\ \vdots \\ u_{l,p} \\ u_{r,0} \\ \vdots \\ u_{r,p} \end{pmatrix}, \tag{6.17}$$

$\underline{\underline{P}}_2$ and $\underline{\underline{P}}_2^T$ perform the correct permutations

$$\begin{aligned} \underline{\underline{P}}_2^T(\underline{\underline{n}} \otimes \underline{\underline{m}})\underline{\underline{P}}_2 &= \underline{\underline{P}}_2^T \begin{pmatrix} -\underline{\underline{m}} & 0 \\ 0 & \underline{\underline{m}} \end{pmatrix} \underline{\underline{P}}_2 = \underline{\underline{P}}_2^T \begin{pmatrix} -m_{0,0} & \dots & -m_{0,p} & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -m_{p,0} & \dots & -m_{p,p} & 0 & \dots & 0 \\ 0 & \dots & 0 & m_{0,0} & \dots & m_{0,p} \\ 0 & \dots & 0 & m_{p,0} & \dots & m_{p,p} \end{pmatrix} \underline{\underline{P}}_2 \\ &= \begin{pmatrix} -m_{0,0} & 0 & \dots & -m_{0,p} & 0 \\ 0 & m_{0,0} & \dots & 0 & m_{0,p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -m_{p,0} & 0 & \dots & -m_{p,p} & 0 \\ 0 & m_{p,0} & \dots & 0 & m_{p,p} \end{pmatrix} = \underline{\underline{m}} \otimes \underline{\underline{n}}. \end{aligned} \tag{6.18}$$

Thus, the one-dimensional SBP property extends directly to the two-dimensional (and also to the d-dimensional) tensor product structure:

Theorem 6.1. *If a one-dimensional SBP method, given by*

- *the symmetric, positive-definite inner product matrix* $\underline{\underline{m}}$,
- *the restriction matrix* $\underline{\underline{r}}$,
- *the boundary integral matrix* $\underline{\underline{b}} = \underline{\underline{I}}_2$,
- *the outer normal matrix* $\underline{\underline{n}} = \mathrm{diag}(-1, 1)$, *and*
- *the differentiation matrix* $\underline{\underline{d}}$,

is extended to two (d) dimensions via tensor product, given by the matrices in equation (6.9)*, the multi-dimensional SBP property* (6.4) *(or* (6.7)*) is satisfied.*

6.2. Linear stability and conservation

In this section, the d-dimensional linear advection equation with constant velocity

$$\partial_t u + \sum_{i=1}^{d} \partial_i u = 0. \tag{6.19}$$

is considered. An SBP CPR method can be formulated as

$$\partial_t \underline{u} + \underline{\underline{D}}\,\underline{f} + \underline{\underline{C}}\,(\underline{f}^{\text{num}} - \underline{\underline{N}}\,(\underline{\underline{I}}_d \otimes \underline{\underline{R}})\underline{f}) = 0, \quad \underline{f} = \begin{pmatrix} \underline{u} \\ \vdots \\ \underline{u} \end{pmatrix}, \tag{6.20}$$

where $\underline{\underline{D}} = \left(\underline{\underline{D}}_1, \ldots, \underline{\underline{D}}_d\right)$ is the divergence operator and $\underline{\underline{N}} = \left(\underline{\underline{N}}_1, \ldots, \underline{\underline{N}}_d\right)$ represents multiplication with the outer normal. Scalar multiplication with $\underline{v}$ and application of the SBP property (6.4) yields

$$\begin{aligned} \underline{v}^T \underline{\underline{M}}\,\partial_t \underline{u} &= -\underline{v}^T \underline{\underline{M}}\,\underline{\underline{D}}\,\underline{f} - \underline{v}^T \underline{\underline{M}}\,\underline{\underline{C}}\,\underline{f}^{\text{num}} + \underline{v}^T \underline{\underline{M}}\,\underline{\underline{C}}\,\underline{\underline{N}}\,(\underline{\underline{I}}_d \otimes \underline{\underline{R}})\underline{f} \\ &= +\underline{v}^T \left(\underline{\underline{D}}_1{}^T, \ldots, \underline{\underline{D}}_d{}^T\right)(\underline{\underline{I}}_d \otimes \underline{\underline{M}})\underline{f} - \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{\underline{N}}\,(\underline{\underline{I}}_d \otimes \underline{\underline{R}})\underline{f} \\ &\quad - \underline{v}^T \underline{\underline{M}}\,\underline{\underline{C}}\,\underline{f}^{\text{num}} + \underline{v}^T \underline{\underline{M}}\,\underline{\underline{C}}\,\underline{\underline{N}}\,(\underline{\underline{I}}_d \otimes \underline{\underline{R}})\underline{f}. \end{aligned} \tag{6.21}$$

Thus, setting $\underline{v} = \underline{1}$, requiring $\underline{1}^T \underline{\underline{M}}\,\underline{\underline{C}} = \underline{1}^T \underline{\underline{R}}^T \underline{\underline{B}}$ and inserting exact differentiation for a constant function results in

$$\frac{\mathrm{d}}{\mathrm{d}t} \underline{1}^T \underline{\underline{M}}\,\underline{u} = -\underline{1}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{f}^{\text{num}}. \tag{6.22}$$

Therefore, since $\underline{f}^{\text{num}}$ is a common numerical flux multiplied with the outer normal, the contributions of two adjacent cells in a conforming grid (of cells with linear coordinates) cancel each other and the scheme is conservative.

Likewise, setting $\underline{v} = \underline{u}$ and using the canonical correction matrix $\underline{\underline{C}} = \underline{\underline{M}}^{-1} \underline{\underline{R}}^T \underline{\underline{B}}$ gives

$$\begin{aligned} \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \|u\|_M^2 &= +\underline{u}^T \left(\underline{\underline{D}}_1{}^T, \ldots, \underline{\underline{D}}_d{}^T\right)(\underline{\underline{I}}_d \otimes \underline{\underline{M}})\underline{f} - \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{\underline{N}}\,(\underline{\underline{I}}_d \otimes \underline{\underline{R}})\underline{f} \\ &\quad - \underline{u}^T \underline{\underline{M}}\,\underline{\underline{C}}\,\underline{f}^{\text{num}} + \underline{u}^T \underline{\underline{M}}\,\underline{\underline{C}}\,\underline{\underline{N}}\,(\underline{\underline{I}}_d \otimes \underline{\underline{R}})\underline{f}. \end{aligned} \tag{6.23}$$

Inserting the flux $\underline{f}$, this can be rewritten as

$$\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \|u\|_M^2 = \sum_{i=1}^{d} \left(\underline{u}^T \underline{\underline{D}}_i{}^T \underline{\underline{M}}\,\underline{u}\right) - \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{f}^{\text{num}}. \tag{6.24}$$

Splitting the sum in two equal parts and applying the SBP property (6.7), this is replaced by

$$\begin{aligned} \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \|u\|_M^2 &= \sum_{i=1}^{d} \left(\frac{1}{2} \underline{u}^T \underline{\underline{D}}_i{}^T \underline{\underline{M}}\,\underline{u} + \frac{1}{2} \underline{u}^T \underline{\underline{D}}_i{}^T \underline{\underline{M}}\,\underline{u}\right) - \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{f}^{\text{num}} \\ &= \sum_{i=1}^{d} \left(\frac{1}{2} \underline{u}^T \underline{\underline{D}}_i{}^T \underline{\underline{M}}\,\underline{u} - \frac{1}{2} \underline{u}^T \underline{\underline{M}}\,\underline{\underline{D}}_i \underline{u} + \frac{1}{2} \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{\underline{N}}_i \underline{\underline{R}}\,\underline{u}\right) - \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{f}^{\text{num}} \\ &= \frac{1}{2} \underline{u}^T \underline{\underline{R}}^T \underline{\underline{B}} \left(\sum_{i=1}^{d} \underline{\underline{N}}_i \underline{\underline{R}}\,\underline{u} - 2\underline{f}^{\text{num}}\right). \end{aligned} \tag{6.25}$$

Thus, if a condition analogous to the one-dimensional investigation for the numerical flux is fulfilled pointwise in a conforming grid, the multi-dimensional scheme is (linearly) stable. Indeed, consider a numerical flux (taking the interpolated values of u from the given cell, the adjacent cell and the outer normal as arguments) of the form

$$f^{\text{num}}(u_-, u_+, n) = \left(\sum_{i=1}^{d} n_i\right) \frac{u_+ + u_-}{2} - \alpha \left|\sum_{i=1}^{d} n_i\right| (u_+ - u_-), \quad \alpha \in [0, 1], \tag{6.26}$$

recovering a central flux for $\alpha = 0$ and a fully upwind flux for $\alpha = 1$. Here, $\sum_{i=1}^{d} n_i$ is the scalar product of the outer normal n and the constant advection velocity $(1, \dots, 1)^T$.

As in the one-dimensional case, an assumption about the boundary basis is necessary for further computations. Mimicking the choice in one space dimension, a nodal basis at the boundary with associated quadrature rule described by a diagonal and positive-definite matrix $\underline{\underline{B}}$ is assumed. In this case, the contribution from the boundary can be computed pointwise at the respective nodes. Using a dense-norm basis at the boundary yields in a sum of terms from different nodes and therefore complicates the investigation. Contrary, assuming a quadrature basis with positive weights at the boundary, the right-hand side of (6.25) can be computed from the contributions of the boundary nodes. Summing over all elements (and using periodic boundary conditions) yields

$$u_- \left[\left(\sum_{i=1}^{d} n_i \right) u_- - 2 f^{\text{num}}(u_-, u_+, n) \right] + u_+ \left[\left(\sum_{i=1}^{d} -n_i \right) u_+ - 2 f^{\text{num}}(u_+, u_-, -n) \right] \tag{6.27}$$

as (a multiple of the) contribution of one node at a boundary. Using the symmetry of the numerical flux, i.e. $f^{\text{num}}(u_+, u_-, -n) = -f^{\text{num}}(u_-, u_+, n)$, this can be rewritten as

$$\begin{aligned} &\left(\sum_{i=1}^{d} n_i \right) (u_-^2 - u_+^2) + 2(u_+ - u_-) f^{\text{num}}(u_-, u_+, n) \\ &= \left(\sum_{i=1}^{d} n_i \right) (u_-^2 - u_+^2) + \left(\sum_{i=1}^{d} n_i \right) (u_+ - u_-)(u_+ + u_-) - 2\alpha \left| \sum_{i=1}^{d} n_i \right| (u_+ - u - -)^2 \\ &= -2\alpha \left| \sum_{i=1}^{d} n_i \right| (u_+ - u - -)^2 \leq 0, \end{aligned} \tag{6.28}$$

since $\alpha \in [0, 1]$. Thus, $\frac{\mathrm{d}}{\mathrm{d}t} \|u\|_M^2 \leq 0$ and the scheme is stable. This proves

Theorem 6.2. *If an SBP CPR method consists of*

- *a nodal basis for the boundary based on a quadrature with positive weights, implying a diagonal and positive-definite* $\underline{\underline{B}}$,
- *the canonical correction matrix* $\underline{\underline{C}} = \underline{\underline{M}}^{-1} \underline{\underline{R}}^T \underline{\underline{B}}$,
- *the numerical flux* (6.26),

and the numerical grid is conforming with boundary operators of adjacent cells projecting on the same nodes, then the scheme (6.20) *for the (constant velocity) linear advection equation* (6.19) *is both conservative and stable in the discrete norm* $\|\cdot\|_M$ *induced by* $\underline{\underline{M}}$.

Similar to section 3.3, other choices of the correction matrix can lead to stable schemes in different norms induced by a matrix $\underline{\underline{M}} + \underline{\underline{K}}$. The results would be similar to those by Castonguay et al. [2012]. Since the computations in two dimensions are very tedious and there does not seem to be a straightforward extension of this idea to nonlinearly stable schemes (see section 4.5), this idea is not further investigated here.

6.3. Stability and conservation for Burgers' equation

As a nonlinear model problem, Burgers' equation in multiple (d) space-dimensions

$$\partial_t u + \sum_{i=1}^{d} \partial_i \frac{u^2}{2} = 0 \tag{6.29}$$

is considered. An SBP CPR method with correction terms for the divergence and restriction can be written as

$$\partial_t \underline{u} + \underline{\underline{D}}\,\underline{f} + \underline{c}_{div} + \underline{\underline{C}}\,(\underline{f}^{\text{num}} - \underline{\underline{N}}\,(\underline{I}_d \otimes \underline{\underline{R}})\underline{f} - \underline{c}_{res}) = 0, \quad \underline{f} = \frac{1}{2}(\underline{I}_d \otimes \underline{u}) \begin{pmatrix} \underline{u} \\ \vdots \\ \underline{u} \end{pmatrix}. \tag{6.30}$$

Inserting the matrices $\underline{\underline{D}} = \left(\underline{\underline{D}}_1, \ldots, \underline{\underline{D}}_d\right)$ and $\underline{\underline{N}} = \left(\underline{\underline{N}}_1, \ldots, \underline{\underline{N}}_d\right)$, representing divergence and multiplication with the outer normal, respectively, this can be rewritten as

$$\partial_t \underline{u} + \frac{1}{2}\sum_{i=1}^{d} \underline{\underline{D}}_i \underline{u}\,\underline{u} + \underline{c}_{div} + \underline{\underline{C}} \left(\underline{f}^{\text{num}} - \frac{1}{2}\sum_{i=1}^{d} \underline{\underline{N}}_i \underline{\underline{R}}\,\underline{u}\,\underline{u} - \underline{c}_{res} \right) = 0. \tag{6.31}$$

Using correction terms similar to the ones in (5.6)

$$\underline{c}_{div} = \frac{1}{3}\sum_{i=1}^{d} \left(\underline{\underline{M}}^{-1} \underline{u}^T \underline{\underline{M}}\,\underline{\underline{D}}_i \underline{u} - \frac{1}{2}\underline{\underline{D}}_i \underline{u}\,\underline{u} \right), \quad \underline{c}_{res} = \frac{1}{6}\sum_{i=1}^{d} \underline{\underline{N}}_i \left((\underline{\underline{R}}\,\underline{u})^2 - \underline{\underline{R}}\,\underline{u}\,\underline{u} \right), \tag{6.32}$$

yields

$$\partial_t \underline{u} = -\frac{1}{3}\sum_{i=1}^{d} \underline{\underline{D}}_i \underline{u}\,\underline{u} - \frac{1}{3}\sum_{i=1}^{d} \underline{\underline{M}}^{-1} \underline{u}^T \underline{\underline{M}}\,\underline{\underline{D}}_i \underline{u} - \underline{\underline{C}} \left(\underline{f}^{\text{num}} - \frac{1}{3}\sum_{i=1}^{d} \underline{\underline{N}}_i \underline{\underline{R}}\,\underline{u}\,\underline{u} - \frac{1}{6}\sum_{i=1}^{d} \underline{\underline{N}}_i (\underline{\underline{R}}\,\underline{u})^2 \right). \tag{6.33}$$

Multiplying with $\underline{v}^T \underline{\underline{M}}$ and inserting the canonical correction matrix $\underline{\underline{C}} = \underline{\underline{M}}^{-1} \underline{\underline{R}}^T \underline{\underline{B}}$ results in

$$\begin{aligned} \underline{v}^T \underline{\underline{M}}\,\partial_t \underline{u} = &-\frac{1}{3}\sum_{i=1}^{d} \underline{v}^T \underline{\underline{M}}\,\underline{\underline{D}}_i \underline{u}\,\underline{u} - \frac{1}{3}\sum_{i=1}^{d} \underline{v}^T \underline{u}^T \underline{\underline{M}}\,\underline{\underline{D}}_i \underline{u} \\ &- \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}} \left(\underline{f}^{\text{num}} - \frac{1}{3}\sum_{i=1}^{d} \underline{\underline{N}}_i \underline{\underline{R}}\,\underline{u}\,\underline{u} - \frac{1}{6}\sum_{i=1}^{d} \underline{\underline{N}}_i (\underline{\underline{R}}\,\underline{u})^2 \right). \end{aligned} \tag{6.34}$$

Application of the SBP property (6.7) gives

$$\begin{aligned} \underline{v}^T \underline{\underline{M}}\,\partial_t \underline{u} = &-\frac{1}{3}\sum_{i=1}^{d} \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{\underline{N}}_i \underline{\underline{R}}\,\underline{u}\,\underline{u} + \frac{1}{3}\sum_{i=1}^{d} \underline{v}^T \underline{\underline{D}}_i{}^T \underline{\underline{M}}\,\underline{u}\,\underline{u} - \frac{1}{3}\sum_{i=1}^{d} \underline{v}^T \underline{u}^T \underline{\underline{M}}\,\underline{\underline{D}}_i \underline{u} \\ &- \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}} \left(\underline{f}^{\text{num}} - \frac{1}{3}\sum_{i=1}^{d} \underline{\underline{N}}_i \underline{\underline{R}}\,\underline{u}\,\underline{u} - \frac{1}{6}\sum_{i=1}^{d} \underline{\underline{N}}_i (\underline{\underline{R}}\,\underline{u})^2 \right) \\ = &\sum_{i=1}^{d} \left(\frac{1}{3}\underline{v}^T \underline{\underline{D}}_i{}^T \underline{\underline{M}}\,\underline{u}\,\underline{u} - \frac{1}{3}\underline{v}^T \underline{u}^T \underline{\underline{M}}\,\underline{\underline{D}}_i \underline{u} + \frac{1}{6}\underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{\underline{N}}_i (\underline{\underline{R}}\,\underline{u})^2 \right) - \underline{v}^T \underline{\underline{R}}^T \underline{\underline{B}}\,\underline{f}^{\text{num}}. \end{aligned} \tag{6.35}$$

Investigating conservation by setting $\underline{v} = \underline{1}$,

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} &= \sum_{i=1}^{d}\left(\frac{1}{3}\underline{1}^T\underline{\underline{D}_i}^T\underline{\underline{M}}\,\underline{\underline{u}}\,\underline{u} - \frac{1}{3}\underline{1}^T\underline{\underline{u}}^T\underline{\underline{M}}\,\underline{\underline{D}_i}\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}_i}(\underline{\underline{R}}\,\underline{u})^2\right) - \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\mathrm{num}} \\ &= \sum_{i=1}^{d}\left(-\frac{1}{3}\underline{u}^T\underline{\underline{M}}\,\underline{\underline{D}_i}\underline{u} + \frac{1}{6}\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}_i}(\underline{\underline{R}}\,\underline{u})^2\right) - \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\mathrm{num}},\end{aligned} \tag{6.36}$$

since $\underline{\underline{D}_i}\underline{1} = 0$ and $\underline{\underline{u}}\,\underline{1} = \underline{u}$. Writing the summands of the first term by the SBP property (6.7) as

$$-\frac{1}{3}\underline{u}^T\underline{\underline{M}}\,\underline{\underline{D}_i}\underline{u} = -\frac{1}{6}\underline{u}^T\underline{\underline{M}}\,\underline{\underline{D}_i}\underline{u} + \frac{1}{6}\underline{u}^T\underline{\underline{D}_i}^T\underline{\underline{M}}\,\underline{u} - \frac{1}{6}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}_i}\underline{\underline{R}}\,\underline{u} = -\frac{1}{6}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}_i}\underline{\underline{R}}\,\underline{u}, \tag{6.37}$$

and using the nodal basis with diagonal $\underline{\underline{B}}$ for the boundary, implying

$$\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}_i}\underline{\underline{R}}\,\underline{u} = \underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}_i}(\underline{\underline{R}}\,\underline{u})^2 \tag{6.38}$$

as in the proof of Theorem 5.1, this reduces to

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^T\underline{\underline{M}}\,\underline{u} = -\underline{1}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\mathrm{num}}. \tag{6.39}$$

Thus, the scheme is conservative.

To prove stability, $\underline{v} = \underline{u}$ is inserted to give

$$\begin{aligned}\frac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\|u\|_M^2 &= \sum_{i=1}^{d}\left(\frac{1}{3}\underline{u}^T\underline{\underline{D}_i}^T\underline{\underline{M}}\,\underline{\underline{u}}\,\underline{u} - \frac{1}{3}\underline{u}^T\underline{\underline{u}}^T\underline{\underline{M}}\,\underline{\underline{D}_i}\underline{u} + \frac{1}{6}\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}_i}(\underline{\underline{R}}\,\underline{u})^2\right) - \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\mathrm{num}} \\ &= \frac{1}{6}\sum_{i=1}^{d}\left(\underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{\underline{N}_i}(\underline{\underline{R}}\,\underline{u})^2\right) - \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\,\underline{f}^{\mathrm{num}} \\ &= \underline{u}^T\underline{\underline{R}}^T\underline{\underline{B}}\left(\frac{1}{6}\sum_{i=1}^{d}\underline{\underline{N}_i}(\underline{\underline{R}}\,\underline{u})^2 - \underline{f}^{\mathrm{num}}\right),\end{aligned} \tag{6.40}$$

similar to the one-dimensional case (4.23). As in the proof proceeding Theorem 6.2, a nodal basis for the boundary corresponding to a quadrature with positive weights and a conforming grid with restriction operators mapping the values of adjacent cells to the same nodes is assumed.

Inserting a local Lax-Friedrichs flux (compare to (4.39))

$$f^{\mathrm{num}}(u_-, u_+, n) = \frac{1}{2}\sum_{i=1}^{d} n_i\left(\frac{u_-^2}{2} + \frac{u_+^2}{2}\right) - \frac{\max\{|u_-|, |u_+|\}}{2}\left|\sum_{i=1}^{d} n_i\right|(u_+ - u_-) \tag{6.41}$$

in the contribution of one node at the boundary yields

$$\begin{aligned}&u_-\left[\frac{1}{6}\sum_{i=1}^{d} n_i u_-^2 - f^{\mathrm{num}}(u_-, u_+, n)\right] + u_+\left[\frac{1}{6}\sum_{i=1}^{d} -n_i u_+^2 - f^{\mathrm{num}}(u_+, u_-, -n)\right] \\ &= \frac{1}{6}\sum_{i=1}^{d} n_i\left(u_-^3 - u_+^3\right) + (u_+ - u_-)f^{\mathrm{num}}(u_-, u_+, n) \\ &= \frac{1}{6}\sum_{i=1}^{d} n_i\left(u_-^3 - u_+^3\right) + \frac{1}{4}\sum_{i=1}^{d} n_i(u_+ - u_-)\left(u_-^2 + u_+^2\right) - \frac{\max\{|u_-|, |u_+|\}}{2}\left|\sum_{i=1}^{d} n_i\right|(u_+ - u_-)^2.\end{aligned} \tag{6.42}$$

This can be rewritten as

$$\begin{aligned}&\frac{1}{12}\left(\sum_{i=1}^{d} n_i\right)\left(u_+^3 - 3u_+^2 u_- + 3u_+ u_-^2 - u_-^3\right) - \frac{\max\{|u_-|,|u_+|\}}{2}\left|\sum_{i=1}^{d} n_i\right|(u_+ - u_-)^2\\ &= \frac{1}{12}\left(\sum_{i=1}^{d} n_i\right)(u_+ - u_-)^2\left(\operatorname{sign}\left(\sum_{i=1}^{d} n_i\right)(u_+ - u_-) - 6\max\{|u_-|,|u_+|\}\right)\end{aligned} \tag{6.43}$$

Thus, since

$$6\max\{|u_-|,|u_+|\} \geq 3(|u_+| + |u_-|) \geq \pm(u_+ - u_-), \tag{6.44}$$

the contribution is non-positive and the resulting scheme therefore stable.

Another possible choice is Osher's flux (see also (4.40))

$$f^{\text{num}}(u_-,u_+,n) = \sum_{i=1}^{d} n_i \cdot \begin{cases} \frac{u_-^2}{2}, & \sum_{i=1}^d n_i u_-, \sum_{i=1}^d n_i u_+ < 0,\\ \frac{u_+^2}{2}, & \sum_{i=1}^d n_i u_-, \sum_{i=1}^d n_i u_+ > 0,\\ \frac{u_+^2}{2} + \frac{u_-^2}{2}, & \sum_{i=1}^d n_i u_- \geq 0 \geq \sum_{i=1}^d n_i u_+,\\ 0, & \sum_{i=1}^d n_i u_- \leq 0 \leq \sum_{i=1}^d n_i u_+. \end{cases} \tag{6.45}$$

Inserting this flux in the contribution of one boundary node for $\sum_{i=1}^d n_i u_-, \sum_{i=1}^d n_i u_+ < 0$ yields

$$\begin{aligned}&u_-\left[\frac{1}{6}\sum_{i=1}^{d} n_i u_-^2 - f^{\text{num}}(u_-,u_+,n)\right] + u_+\left[\frac{1}{6}\sum_{i=1}^{d} -n_i u_+^2 - f^{\text{num}}(u_+,u_-,-n)\right]\\ &= \frac{1}{6}\sum_{i=1}^{d} n_i\left(u_-^3 - u_+^3\right) + (u_+ - u_-) f^{\text{num}}(u_-,u_+,n)\\ &= \frac{1}{6}\sum_{i=1}^{d} n_i\left(u_-^3 - u_+^3\right) + \frac{1}{2}\sum_{i=1}^{d} n_i\,(u_+ - u_-)\,u_-^2\end{aligned} \tag{6.46}$$

This is non-positive, since

$$\begin{aligned}&\frac{1}{6}\sum_{i=1}^{d} n_i\left(-u_+^3 + u_-^3 + 3u_+u_-^2 - 3u_-^3\right) \leq 0\\ \Leftrightarrow &\frac{1}{6}\left(\sum_{i=1}^{d} n_i\right)^3\left(-u_+^3 - 2u_-^3 + 3u_+u_-^2\right) \leq 0\end{aligned} \tag{6.47}$$

and Young's inequality (4.42) yields

$$\begin{aligned}&\left(\sum_{i=1}^{d} n_i\right)^3\left(-u_+^3 - 2u_-^3 + 3u_+u_-^2\right)\\ &\leq \left(\sum_{i=1}^{d} n_i\right)^3\left(-u_+^3 - 2u_-^3\right) + 3\frac{1}{3}\left(\sum_{i=1}^{d} n_i u_-\right)^3 + 3\frac{2}{3}\left(\sum_{i=1}^{d} n_i u_+\right)^3 = 0.\end{aligned} \tag{6.48}$$

The case $\sum_{i=1}^{d} n_i u_-, \sum_{i=1}^{d} n_i u_+ > 0$ is similar. For $\sum_{i=1}^{d} n_i u_- \geq 0 \geq \sum_{i=1}^{d} n_i u_+$, the contribution is

$$\begin{aligned}
&\frac{1}{6}\sum_{i=1}^{d} n_i \left(u_-^3 - u_+^3\right) + (u_+ - u_-) f^{\text{num}}(u_-, u_+, n) \\
&= \frac{1}{6}\sum_{i=1}^{d} n_i \left(u_-^3 - u_+^3\right) + \frac{1}{2}\sum_{i=1}^{d} n_i (u_+ - u_-)\left(u_+^2 + u_-^2\right) \\
&= \frac{1}{6}\sum_{i=1}^{d} n_i \left(-u_+^3 + u_-^3 + 3u_+^3 - 3u_+^2 u_- + 3u_+ u_-^2 - 3u_-^3\right) \\
&= \frac{1}{6}\sum_{i=1}^{d} n_i \left(u_+^3 - u_-^3 + (u_+ - u_-)^3\right).
\end{aligned} \tag{6.49}$$

This is non-positive, since

$$\begin{aligned}
&\frac{1}{6}\sum_{i=1}^{d} n_i \left(u_+^3 - u_-^3 + (u_+ - u_-)^3\right) \leq 0 \\
\Leftrightarrow &\frac{1}{6}\left(\sum_{i=1}^{d} n_i\right)^3 \left(u_+^3 - u_-^3 + (u_+ - u_-)^3\right) \leq 0
\end{aligned} \tag{6.50}$$

and each term is non-positive. The fourth case $\sum_{i=1}^{d} n_i u_- \leq 0 \leq \sum_{i=1}^{d} n_i u_+$ is similar.

Summing up the results, the following Theorem is proved.

Theorem 6.3. *Assume the numerical flux satisfies*

$$u_- \left[\frac{1}{6}\sum_{i=1}^{d} n_i u_-^2 - f^{\text{num}}(u_-, u_+, n)\right] + u_+ \left[\frac{1}{6}\sum_{i=1}^{d} -n_i u_+^2 - f^{\text{num}}(u_+, u_-, -n)\right] \leq 0, \tag{6.51}$$

Then, an SBP CPR method of the form (6.30) *with correction terms* (6.32) *for both divergence and restriction and*

- *a nodal basis for the boundary based on a quadrature with positive weights, implying a diagonal and positive-definite* $\underline{\underline{B}}$,
- *the canonical correction matrix* $\underline{\underline{C}} = \underline{\underline{M}}^{-1}\underline{\underline{R}}^T\underline{\underline{B}}$,
- *a conforming grid with with boundary operators of adjacent cells projecting on the same nodes,*

for the multi-dimensional Burgers' equation (6.29) *is both conservative and stable in the discrete norm* $\|\cdot\|_M$ *induced by* $\underline{\underline{M}}$.

Numerical fluxes fulfilling this condition are inter alia the local Lax-Friedrichs flux (6.41) *and Osher's flux* (6.45).

7 Summary and further research

The aim of this master's thesis was to compare CPR methods with schemes using SBP operators. The resulting embedding of CPR methods into the general framework of SBP operators and SATs is described in chapter 3, leading to a description using simple linear operators and numerical fluxes in a multi-block manner. Conservation, linear stability and symmetries of the correspondent CPR methods have been obtained, extending the results of Vincent et al. [2011b, 2015]. As common for schemes relying on SBP operators, stability results are given in discrete norms, adapted to the chosen basis.

By a special choice of parameters, the DGSEM of Gassner [2013] is embedded in this framework. Based on a skew-symmetric formulation, the results for nodal bases and diagonal norms are extended to Burgers' equation in chapter 4. Introducing additional correction terms, an extension to nodal bases not including boundary points (i.e. Gauß-Legendre points) is presented.

Investigating an extended analytical framework, generalised correction terms for Burgers' equation are introduced in chapter 5, broadening the range of conservative and stable SBP CPR methods, allowing both general nodal and modal bases. These results extend directly to (traditional) SBP methods without a known analytical basis [Fernández et al., 2014a].

An extension to multiple space dimensions not relying on tensor products is presented in chapter 6. This extension is similar to the numerical framework of Hicken et al. [2015] but has been developed independently.

Some open problems have been mentioned in this master's thesis. Considering fully discrete schemes, an explicit Euler step introduces additional terms that have to be considered, see section 3.9. Introducing artificial dissipation can balance the occurring entropy production, but there does not seem to be a straightforward and explicit estimate of the necessary artificial dissipation, required to render fully discrete schemes with SSP methods stable.

An extension of the idea presented by Jameson [2010], allowing additional correction matrices [Vincent et al., 2011b, 2015], to nonlinear conservation laws would be interesting. However, a first analysis in section 4.5 did not lead to positive results.

The numerical setting provides both advantages and disadvantages compared to the analytical setting used in this master's thesis. Further investigations in this direction would be interesting, as mentioned in section 5.4.

First attempts to generalise the ideas of chapters 4 and 5 to prove stability for different systems of conservation laws (shallow water equations and Euler's equations of gas dynamics) were not successful. Therefore, other attempts have to be considered.

Using the ideas of Tadmor [1987, 2003], an entropy stable scheme could be constructed using an entropy conservative scheme and additional artificial dissipation. LeFloch et al. [2002] constructed high-order entropy conservative numerical fluxes for finite difference schemes (on periodic domains). By adding artificial dissipation based on ENO reconstruction procedures, entropy stable schemes have been obtained by Fjordholm et al. [2012], Fjordholm [2012]. The stability is based on

a critical sign property of the reconstruction procedure [Fjordholm et al., 2013, Fjordholm and Ray, 2015].

Fisher et al. [2013] investigated SBP operators and constructed a representation relying on (telescoping) flux differences, even for skew-symmetric formulations. Extending these results and using entropy conservative, two-point numerical fluxes [Tadmor, 2003, Ismail and Roe, 2009] as ingredient, high-order entropy stable schemes on finite domains using SBP operators have been constructed [Fisher and Carpenter, 2013, Carpenter and Fisher, 2013, Carpenter et al., 2014, Parsani et al., 2014, 2015, Carpenter et al., 2015].

Since there are systems of conservation laws not known to be endowed with canonical forms obtained as skew-symmetric formulations allowing entropy estimates as in the case of Burgers' equation (e.g. Euler's equations of gas dynamics), the flux difference formulation is advantageous. However, extensions to multiple dimensions still rely on tensor products. Thus, enabling complex geometries is only possible by using curvilinear coordinates, introducing additional problems due to varying coefficients that have to be balanced by corresponding corrections. For the two-dimensional shallow water equations, Wintermeyer et al. [2015] extended the one-dimensional formulation of Gassner et al. [2016] relying on a skew-symmetric form. However, the new correction terms lead to an inefficient formulation for implementation. Instead, they used a flux difference formulation.

For complex geometries, another desirable option would be a discretisation based on simplex elements instead of curved cubes. However, since Lobatto type cubature rules do not exist on triangles [Xu, 2011], a direct extension of Lobatto-Legendre nodes is not possible. Additionally, first attempts to extend the correction terms for the restriction to the boundary to systems of conservation laws (the shallow water equations and Euler's equations in one space dimension) were not successful. The problem is the impossibility to permute interpolation to the boundary and nonlinear operations and manifests for several choices of skew-symmetric splittings. Even the staggered grid method of Carpenter et al. [2015] using Gauß-Legendre nodes as solution points relies on Lobatto-Legendre nodes for the calculations. Additionally, the Lobatto-Legendre nodes have to correspond to a higher-order discretisation, as can be seen from the incompatibility of the chosen interpolation operators proved by Lundquist and Nordström [2015].

A Some bases

To compute the matrices $\underline{\underline{M}}, \underline{\underline{D}}$ for the nodal bases using Chebyshev points, the associated matrices in a modal Legendre basis are used. The coordinate transformation from a nodal basis with nodes $\xi_0, \ldots, \xi_p$ to a modal basis of Legendre polynomials $\phi_0, \ldots, \phi_p$ of degree $\leq p$ is given by the Vandermonde matrix $\underline{\underline{V}}$ with $V_{i,j} = \phi_j(\xi_i)$. Writing vectors and matrices with regard to the modal basis with $\hat{\cdot}$, the transformation is $\underline{\underline{V}}\,\underline{\hat{u}} = \underline{u}$. Thus, operators like the derivative are transformed as $\underline{\underline{\hat{D}}} = \underline{\underline{V}}^{-1}\underline{\underline{D}}\,\underline{\underline{V}}$ and matrices associated with a scalar product like $\underline{\underline{M}}$ as $\underline{\underline{\hat{M}}} = \underline{\underline{V}}^{T}\underline{\underline{M}}\,\underline{\underline{V}}$.

The Legendre polynomials can be represented by Rodrigues' formula [Abramowitz and Stegun, 1972, equation 8.6.18]

$$\phi_p(x) = \frac{1}{2^p\, p!}\frac{\mathrm{d}^p}{\mathrm{d}x^p}(x^2-1)^p \tag{A.1}$$

and are orthogonal in $L^2[-1,1]$ with $\left\|\phi_p\right\|^2 = 2/(2p+1)$. Their boundary values are $\phi_p(1) = 1$ and $\phi_p(-1) = (-1)^p$. Due to Rodrigues' formula, they are symmetric for even p and antisymmetric for odd p. Additionally, they obey

$$\begin{aligned}
\phi'_{p+1}(x) &= \frac{1}{2^{p+1}\,(p+1)!}\frac{\mathrm{d}^{p+2}}{\mathrm{d}x^{p+2}}(x^2-1)^{p+1} \\
&= \frac{1}{2^{p+1}\,(p+1)!}\frac{\mathrm{d}^{p+1}}{\mathrm{d}x^{p+1}}\left[2(p+1)x(x^2-1)^p\right] \\
&= \frac{1}{2^p\,p!}\frac{\mathrm{d}^p}{\mathrm{d}x^p}\frac{\mathrm{d}}{\mathrm{d}x}\left[x(x^2-1)^p\right] \\
&= \frac{1}{2^p\,p!}\frac{\mathrm{d}^p}{\mathrm{d}x^p}\left[(x^2-1)^p + 2px^2(x^2-1)^{p-1}\right] \\
&= \frac{1}{2^p\,p!}\frac{\mathrm{d}^p}{\mathrm{d}x^p}\left[(2p+1)(x^2-1)^p + 2p(x^2-1)^{p-1}\right] \\
&= (2p+1)\frac{1}{2^p\,p!}\frac{\mathrm{d}^p}{\mathrm{d}x^p}(x^2-1)^p + \frac{1}{2^{p-1}\,(p-1)!}\frac{\mathrm{d}^p}{\mathrm{d}x^p}(x^2-1)^{p-1} \\
&= (2p+1)\phi_p(x) + \phi'_{p-1}(x).
\end{aligned} \tag{A.2}$$

The first three Legendre polynomials are $\phi_0(x) = 1$, $\phi_1(x) = x$, $\phi_2(x) = (3x^2-1)/2$. Therefore, the modal matrices are

$$\underline{\underline{\hat{M}}} = \begin{pmatrix} 2 & & & \\ & \frac{2}{3} & & \\ & & \ddots & \\ & & & \frac{2}{2p+1} \end{pmatrix}, \quad \underline{\underline{\hat{D}}} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 3 & 0 & 3 & \ldots \\ 0 & 0 & 0 & 5 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 7 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{A.3}$$

Using $p = 2$ as an example, the nodal bases with dense norm are given by the following matrices (with 64 bit floating point numbers).

- The roots of the Chebyshev polynomials of first kind are $\xi_i = \cos\left(\frac{2i+1}{2p+2}\pi\right)$, for $i = 0, \ldots, p$. The Vandermonde matrix using 64 bit floating point numbers is approximately

$$\underline{\underline{V}} = \begin{pmatrix} 1.0 & 0.866\,025\,403\,784\,438\,7 & 0.625 \\ 1.0 & 6.123\,233\,995\,736\,766 \times 10^{-17} & -0.5 \\ 1.0 & -0.866\,025\,403\,784\,438\,7 & 0.625 \end{pmatrix}. \tag{A.4}$$

Calculating the mass matrix as $\underline{\underline{M}} = \underline{\underline{V}}^{-T}\underline{\underline{\hat{M}}}\,\underline{\underline{V}}$ results in

$$\underline{\underline{M}} = \begin{pmatrix} 0.399\,999\,999\,999\,999\,9 & 0.088\,888\,888\,888\,888\,71 & -0.044\,444\,444\,444\,444\,45 \\ 0.088\,888\,888\,888\,888\,8 & 0.933\,333\,333\,333\,333\,3 & 0.088\,888\,888\,888\,888\,96 \\ -0.044\,444\,444\,444\,444\,37 & 0.088\,888\,888\,888\,888\,99 & 0.399\,999\,999\,999\,999\,97 \end{pmatrix}. \tag{A.5}$$

The restriction (interpolation to the boundary) and boundary matrices used are

$$\underline{\underline{R}} = \begin{pmatrix} 0.089\,316\,397\,477\,040\,87 & -0.333\,333\,333\,333\,333\,2 & 1.244\,016\,935\,856\,292\,2 \\ 1.244\,016\,935\,856\,292\,2 & -0.333\,333\,333\,333\,333\,2 & 0.089\,316\,397\,477\,040\,82 \end{pmatrix}, \tag{A.6}$$

$$\underline{\underline{B}} = \begin{pmatrix} -1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}. \tag{A.7}$$

Computing the derivative matrix via $\underline{\underline{D}} = \underline{\underline{V}}\,\underline{\underline{\hat{D}}}\,\underline{\underline{V}}^{-1}$ yields

$$\underline{\underline{D}} = \begin{pmatrix} 1.732\,050\,807\,568\,877\,2 & -2.309\,401\,076\,758\,503 & 0.577\,350\,269\,189\,625\,6 \\ 0.577\,350\,269\,189\,625\,7 & -1.632\,862\,398\,863\,137\,5 \times 10^{-16} & -0.577\,350\,269\,189\,625\,6 \\ -0.577\,350\,269\,189\,625\,7 & 2.309\,401\,076\,758\,503 & -1.732\,050\,807\,568\,877\,2 \end{pmatrix}. \tag{A.8}$$

- The extrema of the Chebyshev polynomials of first kind are $\xi_i = \cos\left(\frac{i}{p}\pi\right)$, for $i = 0, \ldots, p$. Thus, the matrices are

$$\underline{\underline{V}} = \begin{pmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 6.123\,233\,995\,736\,766 \times 10^{-17} & -0.5 \\ 1.0 & -1.0 & 1.0 \end{pmatrix}, \tag{A.9}$$

$$\underline{\underline{M}} = \begin{pmatrix} 0.266\,666\,666\,666\,666\,66 & 0.133\,333\,333\,333\,333\,25 & -0.066\,666\,666\,666\,666\,65 \\ 0.133\,333\,333\,333\,333\,3 & 1.066\,666\,666\,666\,666\,4 & 0.133\,333\,333\,333\,333\,47 \\ -0.066\,666\,666\,666\,666\,68 & 0.133\,333\,333\,333\,333\,44 & 0.266\,666\,666\,666\,666\,7 \end{pmatrix}, \tag{A.10}$$

$$\underline{\underline{R}} = \begin{pmatrix} 0.0 & 0.0 & 1.0 \\ 1.0 & 0.0 & 0.0 \end{pmatrix}, \tag{A.11}$$

$$\underline{\underline{D}} = \begin{pmatrix} 1.500\,000\,000\,000\,000\,2 & -2.0 & 0.499\,999\,999\,999\,999\,8 \\ 0.500\,000\,000\,000\,000\,1 & -1.224\,646\,799\,147\,353 \times 10^{-16} & -0.499\,999\,999\,999\,999\,94 \\ -0.500\,000\,000\,000\,000\,2 & 2.0 & -1.499\,999\,999\,999\,999\,8 \end{pmatrix}. \tag{A.12}$$

- Finally, the roots of the Chebyshev polynomials of second kind are $\xi_i = \cos\left(\frac{i+1}{p+2}\pi\right)$, for $i = 0, \ldots, p$. Therefore, the matrices are

$$\underline{\underline{V}} = \begin{pmatrix} 1.0 & 0.707\,106\,781\,186\,547\,6 & 0.250\,000\,000\,000\,000\,1 \\ 1.0 & 6.123\,233\,995\,736\,766 \times 10^{-17} & -0.5 \\ 1.0 & -0.707\,106\,781\,186\,547\,5 & 0.249\,999\,999\,999\,999\,9 \end{pmatrix}, \tag{A.13}$$

$$\underline{\underline{M}} = \begin{pmatrix} 0.733\,333\,333\,333\,333\,2 & -0.133\,333\,333\,333\,333\,3 & 0.066\,666\,666\,666\,666\,79 \\ -0.133\,333\,333\,333\,333\,3 & 0.933\,333\,333\,333\,333\,2 & -0.133\,333\,333\,333\,333\,33 \\ 0.066\,666\,666\,666\,666\,79 & -0.133\,333\,333\,333\,333\,33 & 0.733\,333\,333\,333\,333\,3 \end{pmatrix}, \tag{A.14}$$

$$\underline{\underline{R}} = \begin{pmatrix} 0.292\,893\,218\,813\,452\,6 & -1.000\,000\,000\,000\,000\,2 & 1.707\,106\,781\,186\,547\,7 \\ 1.707\,106\,781\,186\,547\,7 & -0.999\,999\,999\,999\,999\,9 & 0.292\,893\,218\,813\,452\,4 \end{pmatrix}, \tag{A.15}$$

$$\underline{\underline{D}} = \begin{pmatrix} 2.121\,320\,343\,559\,643 & -2.828\,427\,124\,746\,19 & 0.707\,106\,781\,186\,547\,2 \\ 0.707\,106\,781\,186\,547\,5 & -3.558\,369\,867\,163\,396 \times 10^{-17} & -0.707\,106\,781\,186\,547\,5 \\ -0.707\,106\,781\,186\,547\,9 & 2.828\,427\,124\,746\,19 & -2.121\,320\,343\,559\,642 \end{pmatrix}. \tag{A.16}$$

Additionally, the diagonal-norm nodal bases are

- Gauß-Legendre basis (i.e. $p+1$ nodes and weights yielding an exact quadrature for polynomials of degree $\leq 2p+1$) with matrices

$$\underline{\underline{M}} = \begin{pmatrix} 0.555\,555\,555\,555\,555\,4 & 0.0 & 0.0 \\ 0.0 & 0.888\,888\,888\,888\,888\,8 & 0.0 \\ 0.0 & 0.0 & 0.555\,555\,555\,555\,555\,4 \end{pmatrix}, \tag{A.17}$$

$$\underline{\underline{R}} = \begin{pmatrix} 1.478\,830\,557\,701\,236\,2 & -0.666\,666\,666\,666\,666\,5 & 0.187\,836\,108\,965\,430\,5 \\ 0.187\,836\,108\,965\,430\,5 & -0.666\,666\,666\,666\,666\,4 & 1.478\,830\,557\,701\,236 \end{pmatrix}, \tag{A.18}$$

$$\underline{\underline{D}} = \begin{pmatrix} -1.936\,491\,673\,103\,709 & 2.581\,988\,897\,471\,611\,6 & -0.645\,497\,224\,367\,902\,8 \\ -0.645\,497\,224\,367\,902\,6 & -2.465\,190\,328\,815\,662 \times 10^{-31} & 0.645\,497\,224\,367\,902\,6 \\ 0.645\,497\,224\,367\,902\,8 & -2.581\,988\,897\,471\,611\,6 & 1.936\,491\,673\,103\,709 \end{pmatrix}. \tag{A.19}$$

- Lobatto-Legendre basis (i.e. $p+1$ weights and nodes with both boundary nodes yielding an exact quadrature for polynomials of degree $\leq 2p-1$) with matrices

$$\underline{\underline{M}} = \begin{pmatrix} 0.333\,333\,333\,333\,333\,3 & 0.0 & 0.0 \\ 0.0 & 1.333\,333\,333\,333\,333\,3 & 0.0 \\ 0.0 & 0.0 & 0.333\,333\,333\,333\,333\,3 \end{pmatrix}, \tag{A.20}$$

$$\underline{\underline{R}} = \begin{pmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{pmatrix}, \tag{A.21}$$

$$\underline{\underline{D}} = \begin{pmatrix} -1.5 & 2.0 & -0.5 \\ -0.5 & 0.0 & 0.5 \\ 0.5 & -2.0 & 1.5 \end{pmatrix}. \tag{A.22}$$

Parts of this appendix have been published by order of Professor Sonar [Ranocha et al., 2015a].

Bibliography

Q. Abbas, E. van der Weide, and J. Nordström. Accurate and stable calculations involving shocks using a new hybrid scheme. In *19th AIAA Computational Fluid Dynamics Conference*. American Institute of Aeronautics and Astronautics, 2009.

Q. Abbas, E. van der Weide, and J. Nordström. Energy stability of the MUSCL scheme. In *Numerical Mathematics and Advanced Applications 2009*, pages 61–68. Springer, 2010.

M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions*. National Bureau of Standards, 1972.

Y. Allaneau and A. Jameson. Connections between the filtered discontinuous Galerkin method and the flux reconstruction approach to high order discretizations. *Computer Methods in Applied Mechanics and Engineering*, 200(49):3628–3636, 2011.

K. Asthana and A. Jameson. High-order flux reconstruction schemes with minimal dispersion and dissipation. *Journal of Scientific Computing*, 62(3):913–944, 2015.

M. H. Carpenter and T. C. Fisher. High-order entropy stable formulations for computational fluid dynamics. In *21st AIAA Computational Fluid Dynamics Conference*. American Institute of Aeronautics and Astronautics, 2013.

M. H. Carpenter, D. Gottlieb, and S. Abarbanel. Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: Methodology and application to high-order compact schemes. *Journal of Computational Physics*, 111(2):220–236, 1994.

M. H. Carpenter, T. C. Fisher, E. J. Nielsen, and S. H. Frankel. Entropy stable spectral collocation schemes for the Navier-Stokes equations: Discontinuous interfaces. *SIAM Journal on Scientific Computing*, 36(5):B835–B867, 2014.

M. H. Carpenter, M. Parsani, T. C. Fisher, and E. J. Nielsen. Entropy stable staggered grid spectral collocation for the Burgers' and compressible Navier-Stokes equations. Technical Report NASA/TM-2015-218990, NASA, NASA Langley Research Center, Hampton, VA 23681-2199, United States, December 2015.

P. Castonguay, P. E. Vincent, and A. Jameson. A new class of high-order energy stable flux reconstruction schemes for triangular elements. *Journal of Scientific Computing*, 51(1):224–256, 2012.

D. De Grazia, G. Mengaldo, D. Moxey, P. E. Vincent, and S. Sherwin. Connections between the discontinuous Galerkin method and high-order flux reconstruction schemes. *International journal for numerical methods in fluids*, 75(12):860–877, 2014.

S. Eriksson, Q. Abbas, and J. Nordström. A stable and conservative method for locally adapting the design order of finite difference schemes. *Journal of Computational Physics*, 230(11):4216–4231, 2011.

D. C. D. R. Fernández and D. W. Zingg. New diagonal-norm summation-by-parts operators for the first derivative with increased order of accuracy. In *22nd AIAA Computational Fluid Dynamics Conference*. American Institute of Aeronautics and Astronautics, 2015.

D. C. D. R. Fernández, P. D. Boom, and D. W. Zingg. A generalized framework for nodal first derivative summation-by-parts operators. *Journal of Computational Physics*, 266:214–239, 2014a.

D. C. D. R. Fernández, J. E. Hicken, and D. W. Zingg. Review of summation-by-parts operators with simultaneous approximation terms for the numerical solution of partial differential equations. *Computers & Fluids*, 95:171–196, 2014b.

T. C. Fisher and M. H. Carpenter. High-order entropy stable finite difference schemes for nonlinear conservation laws: Finite domains. Technical Report NASA/TM-2013-217971, NASA, NASA Langley Research Center, Hampton, VA 23681-2199, United States, February 2013.

T. C. Fisher, M. H. Carpenter, J. Nordström, N. K. Yamaleev, and C. Swanson. Discretely conservative finite-difference formulations for nonlinear conservation laws in split form: Theory and boundary conditions. *Journal of Computational Physics*, 234:353–375, 2013.

U. S. Fjordholm. *High-order accurate entropy stable numerical schemes for hyperbolic conservation laws*. PhD thesis, Eidgenössische Technische Hochschule ETH Zürich, 2012.

U. S. Fjordholm and D. Ray. A sign preserving WENO reconstruction method. *Journal of Scientific Computing*, pages 1–22, 2015.

U. S. Fjordholm, S. Mishra, and E. Tadmor. Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws. *SIAM Journal on Numerical Analysis*, 50 (2):544–573, 2012.

U. S. Fjordholm, S. Mishra, and E. Tadmor. ENO reconstruction and ENO interpolation are stable. *Foundations of Computational Mathematics*, 13(2):139–159, 2013.

G. J. Gassner. A skew-symmetric discontinuous Galerkin spectral element discretization and its relation to SBP-SAT finite difference methods. *SIAM Journal on Scientific Computing*, 35(3):A1233–A1253, 2013.

G. J. Gassner. A kinetic energy preserving nodal discontinuous Galerkin spectral element method. *International Journal for Numerical Methods in Fluids*, 76(1):28–50, 2014.

G. J. Gassner and D. A. Kopriva. A comparison of the dispersion and dissipation errors of Gauss and Gauss-Lobatto discontinuous Galerkin spectral element methods. *SIAM Journal on Scientific Computing*, 33(5):2560–2579, 2011.

G. J. Gassner, A. R. Winters, and D. A. Kopriva. A well balanced and entropy conservative discontinuous Galerkin spectral element method for the shallow water equations. *Applied Mathematics and Computation*, 272:291–308, 2016.

J. E. Hicken and D. W. Zingg. Summation-by-parts operators and high-order quadrature. *Journal of Computational and Applied Mathematics*, 237(1):111–125, 2013.

J. E. Hicken, D. C. D. R. Fernández, and D. W. Zingg. Multidimensional summation-by-parts operators: General theory and application to simplex elements. *arXiv preprint arXiv:1505.03125*, 2015.

H. Huynh. A flux reconstruction approach to high-order schemes including discontinuous Galerkin methods. *AIAA paper*, 4079:2007, 2007.

H. Huynh, Z. J. Wang, and P. E. Vincent. High-order methods for computational fluid dynamics: A brief review of compact differential formulations on unstructured grids. *Computers & Fluids*, 98: 209–220, 2014.

F. Ismail and P. L. Roe. Affordable, entropy-consistent Euler flux functions II: Entropy production at shocks. *Journal of Computational Physics*, 228(15):5410–5436, 2009.

A. Jameson. A proof of the stability of the spectral difference method for all orders of accuracy. *Journal of Scientific Computing*, 45(1-3):348–358, 2010.

A. Jameson, P. E. Vincent, and P. Castonguay. On the non-linear stability of flux reconstruction schemes. *Journal of Scientific Computing*, 50(2):434–445, 2012.

D. A. Kopriva and G. J. Gassner. On the quadrature and weak form choices in collocation type discontinuous Galerkin spectral element methods. *Journal of Scientific Computing*, 44(2):136–155, 2010.

D. A. Kopriva and G. J. Gassner. An energy stable discontinuous Galerkin spectral element discretization for variable coefficient advection problems. *SIAM Journal on Scientific Computing*, 36(4): A2076–A2099, 2014.

J. E. Kozdon and L. C. Wilcox. Stable coupling of nonconforming, high-order finite difference methods. *arXiv preprint arXiv:1410.5746v3*, 2015.

H.-O. Kreiss and G. Scherer. Finite element and finite difference methods for hyperbolic partial differential equations. *Mathematical aspects of finite elements in partial differential equations*, (33):195–212, 1974.

P. G. LeFloch, J.-M. Mercier, and C. Rohde. Fully discrete, entropy conservative schemes of arbitrary order. *SIAM Journal on Numerical Analysis*, 40(5):1968–1992, 2002.

T. Lundquist and J. Nordström. On the suboptimal accuracy of summation-by-parts schemes with non-conforming block interfaces. 2015. Submitted to Journal of Computational Physics.

K. Mattsson. Boundary procedures for summation-by-parts operators. *Journal of Scientific Computing*, 18(1):133–153, 2003.

K. Mattsson and M. Almquist. A solution to the stability issues with block norm summation by parts operators. *Journal of Computational Physics*, 253:418–442, 2013.

K. Mattsson and M. H. Carpenter. Stable and accurate interpolation operators for high-order multiblock finite difference methods. *SIAM Journal on Scientific Computing*, 32(4):2298–2320, 2010.

K. Mattsson, M. Svärd, and J. Nordström. Stable and accurate artificial dissipation. *Journal of Scientific Computing*, 21(1):57–79, 2004.

K. Mattsson, M. Almquist, and M. H. Carpenter. Optimal diagonal-norm SBP operators. *Journal of Computational Physics*, 264:91–111, 2014.

A. Nissen, K. Kormann, M. Grandin, and K. Virta. Stable difference methods for block-oriented adaptive grids. *Journal of Scientific Computing*, 65(2):486–511, 2015.

J. Nordström. Conservative finite difference formulations, variable coefficients, energy estimates and artificial dissipation. *Journal of Scientific Computing*, 29(3):375–404, 2006.

J. Nordström and P. Eliasson. New developments for increased performance of the SBP-SAT finite difference technique. In *IDIHOM: Industrialization of High-Order Methods-A Top-Down Approach*, pages 467–488. Springer, 2015.

M. Parsani, M. H. Carpenter, and E. J. Nielsen. Entropy stable wall boundary conditions for compressible Navier-Stokes equations. Technical Report NASA/TM-2014-218282, NASA, NASA Langley Research Center, Hampton, VA 23681-2199, United States, June 2014.

M. Parsani, M. H. Carpenter, and E. J. Nielsen. Entropy stable discontinuous interfaces coupling for the three-dimensional compressible Navier-Stokes equations. *Journal of Computational Physics*, 290:132–138, 2015.

H. Ranocha, P. Öffner, and T. Sonar. Extended skew-symmetric form for summation-by-parts operators. *arXiv preprint arXiv:1511.08408*, 2015a. Submitted to Mathematics of Computation.

H. Ranocha, P. Öffner, and T. Sonar. Summation-by-parts operators for correction procedure via reconstruction. *arXiv preprint arXiv:1511.02052*, 2015b. Accepted by Journal of Computational Physics.

H. Ranocha, P. Öffner, and T. Sonar. Summation-by-parts operators for correction procedure via reconstruction. *Journal of Computational Physics*, 311:299–328, 2016. See also arXiv preprint arXiv:1511.02052 [Ranocha et al., 2015b].

M. Svärd. On coordinate transformations for summation-by-parts operators. *Journal of Scientific Computing*, 20(1):29–42, 2004.

M. Svärd and J. Nordström. Review of summation-by-parts schemes for initial-boundary-value problems. *Journal of Computational Physics*, 268:17–38, 2014.

E. Tadmor. The numerical viscosity of entropy stable schemes for systems of conservation laws. I. *Mathematics of Computation*, 49(179):91–103, 1987.

E. Tadmor. Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems. *Acta Numerica*, 12:451–512, 2003.

E. F. Toro. *Riemann solvers and numerical methods for fluid dynamics: A practical introduction*. Springer Science & Business Media, 2009.

P. E. Vincent, P. Castonguay, and A. Jameson. Insights from von Neumann analysis of high-order flux reconstruction schemes. *Journal of Computational Physics*, 230(22):8134–8154, 2011a.

P. E. Vincent, P. Castonguay, and A. Jameson. A new class of high-order energy stable flux reconstruction schemes. *Journal of Scientific Computing*, 47(1):50–72, 2011b.

P. E. Vincent, A. M. Farrington, F. D. Witherden, and A. Jameson. An extended range of stable-symmetric-conservative flux reconstruction correction functions. *Computer Methods in Applied Mechanics and Engineering*, 296:248–272, 2015.

Z. Wang and H. Gao. A unifying lifting collocation penalty formulation including the discontinuous Galerkin, spectral volume/difference methods for conservation laws on mixed grids. *Journal of Computational Physics*, 228(21):8161–8186, 2009.

N. Wintermeyer, A. R. Winters, G. J. Gassner, and D. A. Kopriva. An entropy stable nodal discontinuous Galerkin method for the two dimensional shallow water equations on unstructured curvilinear meshes with discontinuous bathymetry. *arXiv preprint arXiv:1509.07096*, 2015.

F. D. Witherden and P. E. Vincent. An analysis of solution point coordinates for flux reconstruction schemes on triangular elements. *Journal of Scientific Computing*, 61(2):398–423, 2014.

F. D. Witherden and P. E. Vincent. On the identification of symmetric quadrature rules for finite element methods. *Computers & Mathematics with Applications*, 69(10):1232–1241, 2015.

Y. Xu. On Gauss-Lobatto integration on the triangle. *SIAM Journal on Numerical Analysis*, 49(2): 541–548, 2011.

M. Yu and Z. Wang. On the connection between the correction and weighting functions in the correction procedure via reconstruction method. *Journal of Scientific Computing*, 54(1):227–244, 2013.

Declaration

I declare that this master's thesis has been composed completely by myself and that the work contained herein is my own except where explicitly stated otherwise in the text.
I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.
Parts of this work have been published by order of Professor Sonar [Ranocha et al., 2015b,a, 2016].

Hendrik Ranocha
Braunschweig, February 17, 2016